CLIENT APPLICATION OF caBENCH-TO-BEDSIDE V3.1

End User Manual

Document Change History

Version Number	Date	Contributor	Description
V1.0	July 12, 2007	Washington University/ Persistent Systems	Reviewed
V2.0	June 9, 2008	Washington University/ Persistent Systems	Reviewed
V2.1	December 4, 2008	Washington University/ Persistent Systems	Reviewed
V2.1	December 23, 2008	Washington University/ Persistent Systems	Reviewed by Documentation and Training workspace
V3.0	June 17,2009	Washington University/ Persistent Systems	Reviewed
V3.0	July 24,2009	Washington University/ Persistent Systems	Addressed comments by Documentation and Training workspace
V 3.1	September 10, 2009	Washington University/ Persistent Systems	Draft
V 3.1	September 12, 2009	Washington University/ Persistent Systems	Review
V3.1	September 16, 2009	Washington University Persistent Systems	Reviewed by Documentation and Training workspace

caBIG®
cancer Biomedical Informatics Grid®

This is a U.S. Government work.

September 23, 2009

Development Teams		
Development	***End User Manual***	***Program Management***
Chandrakant Talele[1]	Chandrakant Talele[1]	Dr. Rakesh Nagarajan[2]
Chetan Patil[1]	Madhumita Shrikhande[1]	Srikanth Adiga[1, 2]
Chetan Pundhir[1]	Pooja Arora[1]	
Deepak Shingan[1]	Srikanth Adiga[1,2]	
Gaurav Mehta[1]	Mukesh Sharma[2]	
Pallavi Mistry[1]		
Rajesh Vyas[1]		
[1]Persistent Systems Limited	[2]Washington University	

Contacts and Support	
Training contact	cab2bsupport@bmi.wustl.edu
Customer support contact	cab2bsupport@bmi.wustl.edu

Copyright and License Information

Copyright Notice. Copyright 2005 "School of Medicine, Washington University in St. Louis" ("caBIG® Participant"). cancer Bench-to-Bedside was created with NCI funding and is part of the caBIG® initiative. The software subject to this notice and license includes both human readable source code form and machine readable, binary, object code form (the "caBIG® Software").

This caBIG® Software License (the "License") is between caBIG® Participant and You. "You (or "Your") shall mean a person or an entity, and all other entities that control, are controlled by, or are under common control with the entity. "Control" for purposes of this definition means (i) the direct or indirect power to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

License. Provided that You agree to the conditions described below, caBIG® Participant grants You a non-exclusive, worldwide, perpetual, fully-paid-up, no-charge, irrevocable, transferable and royalty-free right and license in its rights in the caBIG® Software, including any copyright or patent rights therein that may be infringed by the making, using, selling, offering for sale, or importing of caBIG® Software, to (i) use, install, access, operate, execute, reproduce, copy, modify, translate, market, publicly display, publicly perform, and prepare derivative works of the caBIG® Software; (ii) make, have made, use, practice, sell, and offer for sale, and/or otherwise dispose of caBIG® Software (or portions thereof); (iii) distribute and have distributed to and by third parties the caBIG® Software and any modifications and derivative works thereof; and (iv) sublicense the foregoing rights set out in (i), (ii) and (iii) to third parties, including the right to license such rights to further third parties. For sake of clarity, and not by way of limitation, caBIG® Participant shall have no right of accounting or right of payment from You or Your sublicensees for the rights granted under this License. This License is granted at no charge to You. Your downloading, copying, modifying, displaying, distributing or use of caBIG® Software constitutes acceptance of all of the terms and conditions of this Agreement. If you do not agree to such terms and conditions, you have no right to download, copy, modify, display, distribute or use the caBIG® Software.

1. Your redistributions of the source code for the caBIG® Software must retain the above copyright notice, this list of conditions and the disclaimer and limitation of liability of Article 6 below. Your redistributions in object code form must reproduce the above copyright notice, this list of conditions and the disclaimer of Article 6 in the documentation and/or other materials provided with the distribution, if any.

2. Your end-user documentation included with the redistribution, if any, must include the following acknowledgment: "This product includes software developed by School of Medicine, Washington University in St. Louis." If You do not include such end-user documentation, You shall include this acknowledgment in the caBIG® Software itself, wherever such third-party acknowledgments normally appear.

3. You may not use the names "School of Medicine, Washington University in St. Louis", "The National Cancer Institute", "NCI", "Cancer Bioinformatics Grid" or "caBIG®" to endorse or promote products derived from this caBIG® Software. This License does not authorize You to use any trademarks, service marks, trade names,

logos or product names of either caBIG® Participant, NCI or caBIG®, except as required to comply with the terms of this License.

4. For sake of clarity, and not by way of limitation, You may incorporate this caBIG® Software into Your proprietary programs and into any third party proprietary programs. However, if You incorporate the caBIG® Software into third party proprietary programs, You agree that You are solely responsible for obtaining any permission from such third parties required to incorporate the caBIG® Software into such third party proprietary programs and for informing Your sublicensees, including without limitation Your end-users, of their obligation to secure any required permissions from such third parties before incorporating the caBIG® Software into such third party proprietary software programs. In the event that You fail to obtain such permissions, You agree to indemnify caBIG® Participant for any claims against caBIG® Participant by such third parties, except to the extent prohibited by law, resulting from Your failure to obtain such permissions.

5. For sake of clarity, and not by way of limitation, You may add Your own copyright statement to Your modifications and to the derivative works, and You may provide additional or different license terms and conditions in Your sublicenses of modifications of the caBIG® Software, or any derivative works of the caBIG® Software as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

6. THIS caBIG® SOFTWARE IS PROVIDED "AS IS" AND ANY EXPRESSED OR IMPLIED WARRANTIES (INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE) ARE DISCLAIMED. IN NO EVENT SHALL THE SCHOOL OF MEDICINE, WASHINGTON UNIVERSITY IN ST. LOUIS OR ITS AFFILIATES BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS caBIG® SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Contents

About This Guide 1
Purpose 1
Typical User 1
Topics Covered 1
Text Conventions Used 2

Chapter 1 Introduction 3
Conducting Research using caB2B 3
Getting started with caB2B 4
Starting the Application 4
Types of User 4
Secure and Non-Secure Data 6
Links on Home Page 7
Popular Categories 8
Overriding Administrator-defined Services 9
Understanding Categories and Limits 11
What is a Category? 11
Why do we need categories? 11
What is a Multi-model Category (MMC)? 11
What is a Limit? 12
What is a Query? 12

Chapter 2 Searching Data 13
Search data – The Query Wizard 13
Choose Search Category 14
Advanced Search 15
Define Limit 16
Steps for defining a limit 17
Limit Set – Diagrammatic View 18
DAG Operations 19
Connecting the Limits 20
The Information Panel 23
Reset, Previous and Next Buttons 24
Define Search Results View 24
The Next, Previous, Save Query and Service URL Buttons 25
Configuring Services for a Query 25
DCQL Viewer 28
Save Query 29
Why to Save Query? 29
Types of Saved Queries: 30
How to Save Query? 30
Executing a Saved Query 32
View Search Results 34
Record Details 35
The Next, Previous and Save Query Buttons 40
Group Results by Service URL 40
Viewing and Saving Data List 42
Operations on the Data List 44

Chapter 3 Experiments 46
What are Experiments and Projects? 46
Creating an Experiment 46

Viewing the Existing Experiments 48
Opening an Experiment 49
Working with an Experiment 50
The Operations panel 50
Visualization panel 50
Filtering Data 51
Custom Data Category 54
Visualizing Data 56

Appendix A Examples of category search 60
Text-based Search - For category name information 60
Text-based Search - For category and attribute name and description 61
Concept Code Based Search - For category name information 62
Concept code based search - For attribute name information 63

Appendix B Define Limit: Operators and Values 65

Appendix C Record Details View 67
Administrator defined categories 67

Appendix D Types of Filters 68
Using the Range Filter 68
Using the Pattern Filter 69
Using the Enumerated Values Filter 69

Index 70

About This Guide

This section introduces you to caBench-to-Bedside. Topics in this section include

- *Purpose*
- *Typical User*
- *Topics Covered*
- *Text Conventions Used*

Purpose

cancer **B**ench-to-**B**edside (caB2B) is an application that leverages caBIG® tools in a user-friendly graphical user interface (GUI). caB2B is a tool that provides a platform for querying all caBIG tools. The physician scientist can use caB2B to query any caGrid data service to obtain data, and save this data for further downstream analyses. For example, a physician can look for a specimen of interest from a tissue banking software and a lab scientist can find all the microarray experiments done on the same specimen from a microarray repository.

This manual provides the details of how to use caB2B application to search data sets of interest, create an experiment from data sets, analyze data, and visualize data.

The technical staff can install the caB2B application using the instructions provided in the caB2B installation guide. The Client Application accesses the caB2B server, installed by the technical staff. To start Client Application, refer to the *caB2B Installation Manual's Accessing the caB2B Application* section. It is a java-based application. The application is installed on the user's machine when the user accesses it for the first time.

All the chapters in this user's manual start with an introductory statement followed by a list of topics included in that chapter.

Typical User

A typical user of the caB2B application would be a translational research scientist. The term translational research means research which typically begins at "the bench" with basic research — in which scientists study disease at a molecular or cellular level — then progress to the clinical level, or the patient's "bedside." As this application supports the research at the bench as well as at the bedside, scientists would be the typical users.

The user should have the basic computer skills, like running java applications. The user should also be aware of caBIG and its tools. The user should have knowledge of UML models/domain models of the caBIG tools. Awareness of the domain model is required to create and execute the queries using caB2B.

Topics Covered

- *"Chapter 1: Introduction"* explains the scope and function of caB2B

- *"Chapter 2: Searching Data"* describes process of retrieving data by building metadata based queries
- *"Chapter 3: Experiments"* explains how to create and use virtual experiments using the searched data
- *"Appendix A: Examples of category search"*
- *"Appendix B: Define Limit: Operators and Values"*
- *"Appendix C: Record Details View"*
- *"Appendix D: Types of Filters"*

Text Conventions Used

This section lists the conventions used in this guide. The various typefaces represent interface components, keyboard shortcuts, toolbar buttons, dialog box options, and text that you type.

Convention	***Description***	***Example***
Bold	Highlights names of option buttons, check boxes, drop-down menus, menu commands, command buttons, or icons	Click **Search**.
Underlined Blue Text	Indicates a Web or e-mail address	http://domain.com
text in SMALL CAPS	Indicates a keyboard shortcut	Press ENTER
text in SMALL CAPS + text in SMALL CAPS	Indicates keys that are pressed simultaneously	Press SHIFT + CTRL
Italics	Highlights references to other documents, sections, figures, and tables	See *Figure 4.5*.
Italic boldface monospace type	Represents text that you type	In the **New Subset** text box, enter ***Proprietary Proteins***.
Note:	Highlights information of particular importance	**Note:** This concept is used throughout this document
{ }	Surrounds replaceable items	Replace {last name, first name} with the Principal Investigator's name
	Indicates information of particular interest	You cannot execute MMC queries.
WARNING!	Indicates that you should be particularly aware of the information provided	**WARNING!** You cannot modify path

Chapter 1 Introduction

This chapter explains the scope and function of the cancer Bench to Bedside (caB2B) application. It helps you understand how to use caB2B to satisfy diverse biomedical research needs.

This chapter contains the following topics:

- *Conducting Research using caB2B*
- *Getting started with caB2B*
- *Starting the Application*
- *Understanding Categories and Limits*

Conducting Research using caB2B

The **N**ational **C**ancer **I**nstitute **ca**ncer **B**iomedical **I**nformatics **G**rid® (NCI caBIG®) project creates a common, extensible informatics platform that integrates diverse data types and supports interoperable analytic tools. caBIG® is developing separate applications that will facilitate individual steps involved in micro-array analysis. These applications are also useful to bioinformaticians.

caGrid is the infrastructure for caBIG that helps integrate these applications. caGrid helps to perform investigation involving data integration and analytical services from diverse research communities.

cancer **B**ench-to-**B**edside **(caB2B)** is an application that leverages these tools in a user-friendly graphical user interface (GUI). The physician scientist can use caB2B to perform operations such as:

- **Query any caGrid data service to obtain data**

  The data service can be a single data service, multiple data services, or a combination of the two services using semantically interoperable Common Data Elements (CDEs).

  For example, caB2B allows investigators to query tissue banks at multiple cancer centers, design studies that focus on very specific tumor subtypes and to target less common tumors by pooling bio-specimen resources. This alleviates the problem of small sample size in a study.

- **Collect data and create experiments**

  caB2B also enables investigators to perform novel in silico experiments using archived microarray data sets. These capabilities facilitate identifying genes that are up or down regulated in specific cancers with respect to normals and enable investigators to view the data in the context of biological pathways. Investigators gain further understanding of the complex system of cancer biology by identifying genes that are important to the development and treatment of cancer. These capabilities may lead to effective identification of novel drug targets and improved treatment strategies.

- **Perform analyses** by using various grid-enabled analytical services.

caB2B enables individual users to work on common platforms to access caGrid resources in real time from common data and analytical services.

- **Visualize analysis results** by using charts and dendrograms.

Getting started with caB2B

caB2B provides a platform to bring the laboratory research to the clinical level. It enables information sharing along with the continuum from the scientific bench to the clinical bedside and back. The tool helps in querying various caGrid enabled services. For example, it is possible to query for all the up-regulated and down-regulated genes of a malignant tissue sample by using the services available in caB2B. Thus, it helps bridge the gap between the bench and the bedside. caB2B consist of a caB2B Client Application, caB2B Administrative Module and caB2B Web Application. The Administrative Module allows administrator to configure caB2B Client Application where the administrator can load models from caDSR, configure service URL's create customized categories, create intermodel joins, and curate the paths.

- To configure Client and the Web Application refer to caB2B Administration Manual.
- To start caB2B Application, refer to the Installation Manual.
- To search for data sets, refer to *"Search data – The Query Wizard"*.
- To design experiments, refer to *"Experiments"*.

Starting the Application

Before you start the Client Application, the technical staff should have installed the caB2B application and the administrator should have configured the application using the administrative module. To access the Client Application you need to access the caB2B web page or download the caB2B client setup. For details on installing and starting the caB2B Client Application, refer to *Accessing the caB2B Application* section in the *caB2B Installation Manual*.

Once you install the Client Application using the web-start option, it creates an icon on users desktop. Next time you need to double-click the icon to start the application.

Types of User

The application supports two user roles – Anonymous User and caGrid User. You can search for data without signing in, although only publically available data is available. All such users who are querying for data without signing in are Anonymous Users.

The other type of user is caGrid User. These users are the ones who sign into the application using their production caGrid credentials. Currently, caB2B supports all the production grid users having a Dorian Identity Provider. A user can obtain a caGrid account by accessing the following URL and creating an account (https://cagrid-portal.nci.nih.gov/web/guest/register). When caB2B application starts, you see the caB2B Login page as shown in *Figure 1.1.*

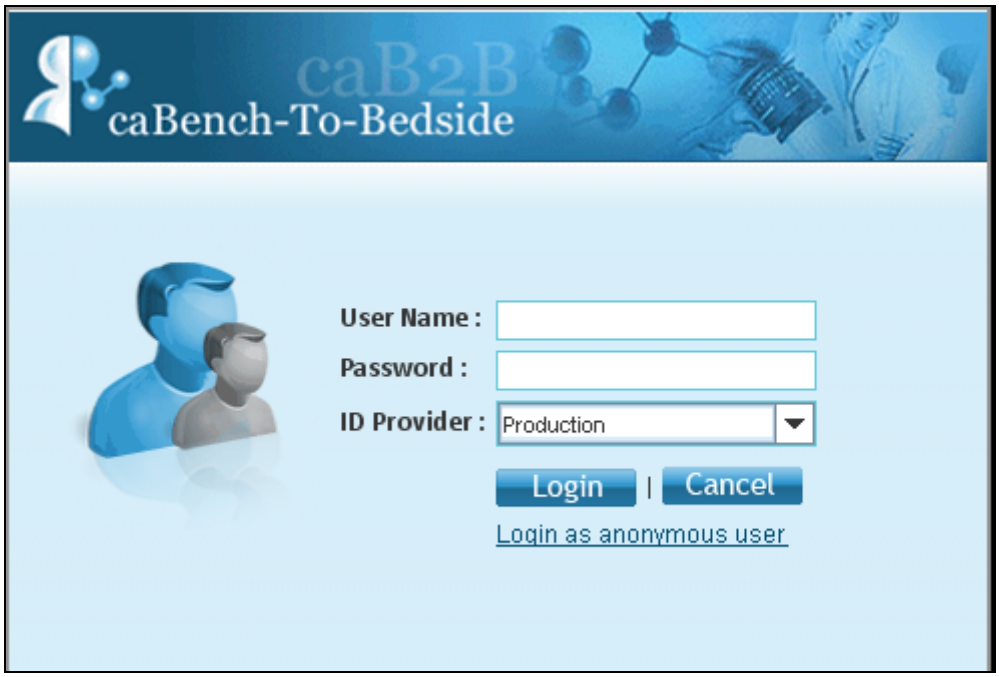


Figure 1.1. caB2B Login Page

To access the Client Application, you can login using your caGrid credential or use application as an anonymous (without grid credentials) user.

To login using your caGrid credentials, follow the steps given below:

1. Type the **User Name** and **Password**.
2. Select an **ID Provider** from the dropdown.
3. Click **Login** to start the application.

Application will authenticate you using entered credentials. On successful authentication, you will see application home page. If authentication is unsuccessful, application will show an error message as shown in *Figure 1.2.*

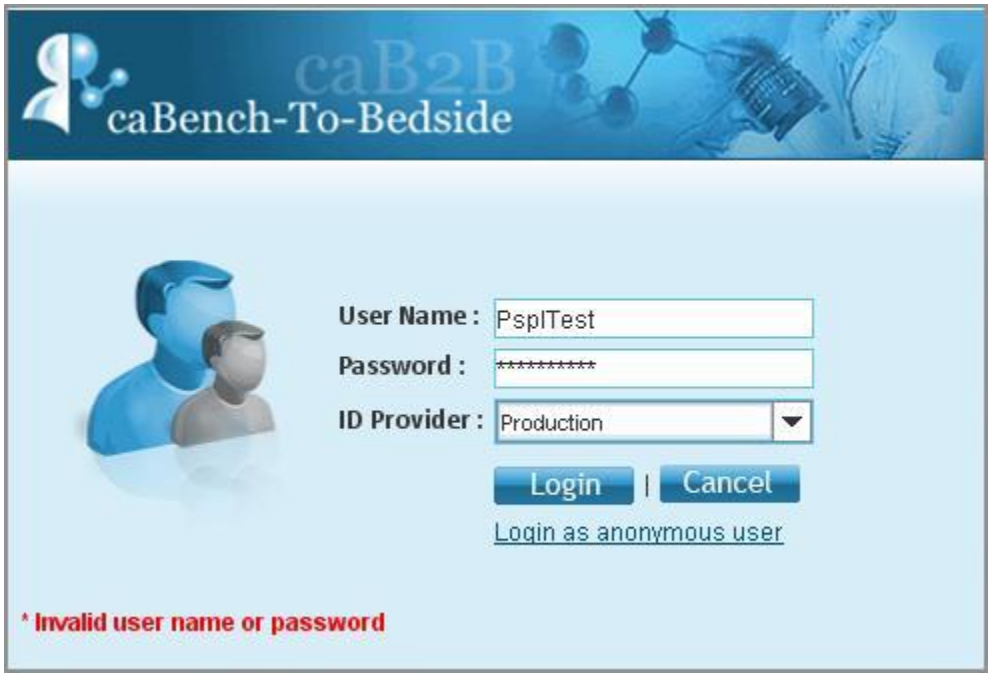


Figure 1.2. Login Error

If you do not have a caGrid account then you can log in as an anonymous user. While using anonymous user login you can only access public data and not the secured data that essentially requires caGrid credentials. It means that you would not be able to see data from any of the secured service like caTissue.

F.Y.I.	The anonymous user's data is not secure. Any user who logs in using the anonymous users link can access it.

Secure and Non-Secure Data

Data on caGrid can be categorized into two types, non-secure data i.e. publically available data and secure data i.e. data that requires authentication for access. Biological data may contain certain protected health information (PHI data). Only an authenticated user should have access to this secure data.

Microarray data available in caArray is an example of non-secure data, whereas biospecimen data available in caTissue is secure data.

When the caB2B application starts, you see the caB2B Home page as shown in *Figure 1.3.*

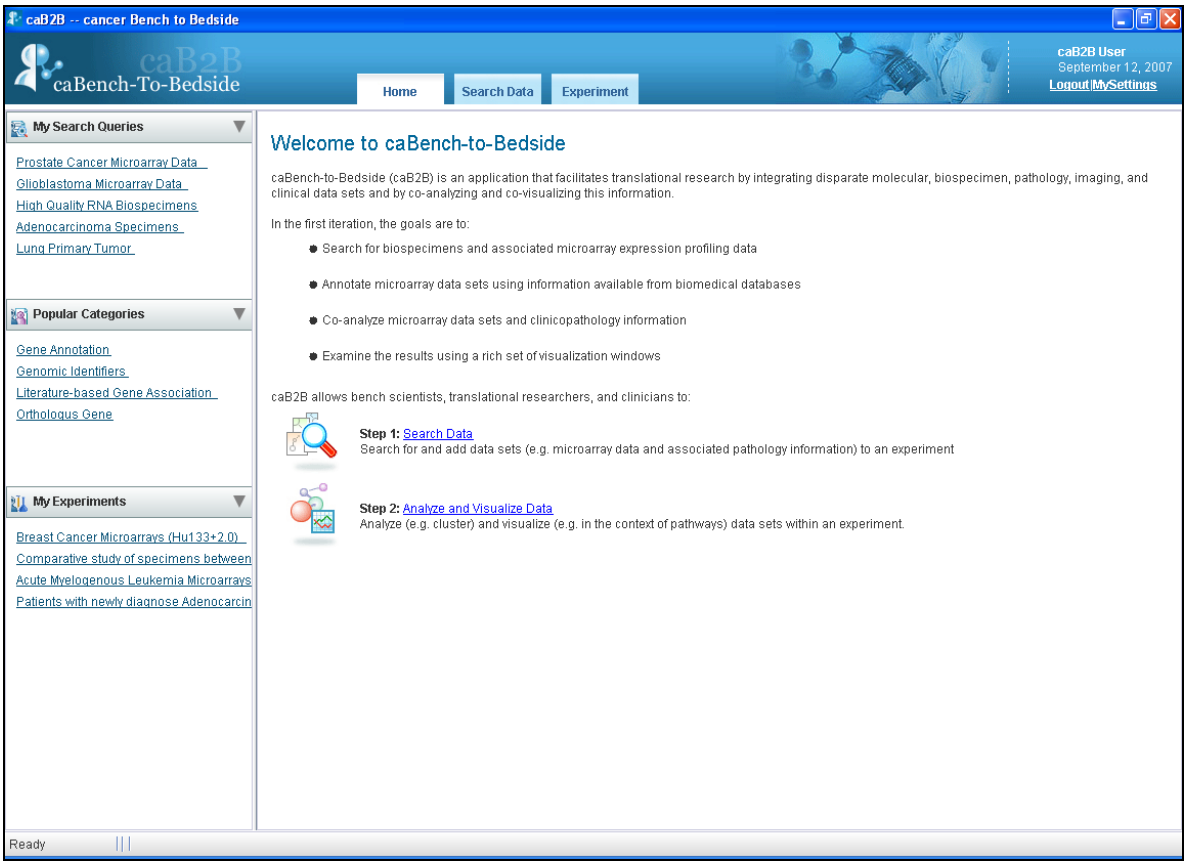


Figure 1.3. caB2B Homepage

Links on Home Page

Client Application's home page provides a quick access to various features available in caB2B. It contains a link to the following features:

- My Search Queries – It gives quick access to all of your saved queries. From the homepage link, you can execute or modify the existing saved query. Only you can see queries created by you. Queries created by other users will not be visible to you.
- Popular Categories – It shows the five most popular categories. Popular categories are the categories that have been queried maximum number of times by the users.
- My Experiments – This panel shows you five most recently created experiments. Only you can see experiments created by you. Experiments created by other users will not be visible to you. You can click on these links to navigate to the experiment.
- Search Data – This link opens the query wizard. You can execute queries and create experiments using this wizard.
- Analyze and Visualize data – This link helps you in analyzing and visualizing the data.

- My Settings – You can override administrator-defined services to define the service instances that you wish to query.
- Logout – Using this you can log out of Client Application

As mentioned in the Conducting Research using caB2B section, you typically perform the following steps:

1. Execute queries to obtain the dataset(s) of interest. The “Searching Data” chapter provides the details for this step.
2. Analyze and visualize the data. The “
3. Experiments” chapter provides the details for this step.

Popular Categories

Popular categories are the categories/classes that are queried the maximum number of times by a user. You can click on any of the categories shown under this and application will navigate to Add Limit page.

The popularity for a class increases when a user queries that particular class. When user queries administrator-defined category then the popularity of the classes, composing that category increases along with the category. For example, if a user queries for a class gene then its popularity would increase by one. Similarly, for administrator-defined category like genomic identifiers consisting of gene, messenger RNA and protein, the popularity of the category genomic identifier and individual classes will increase by one.

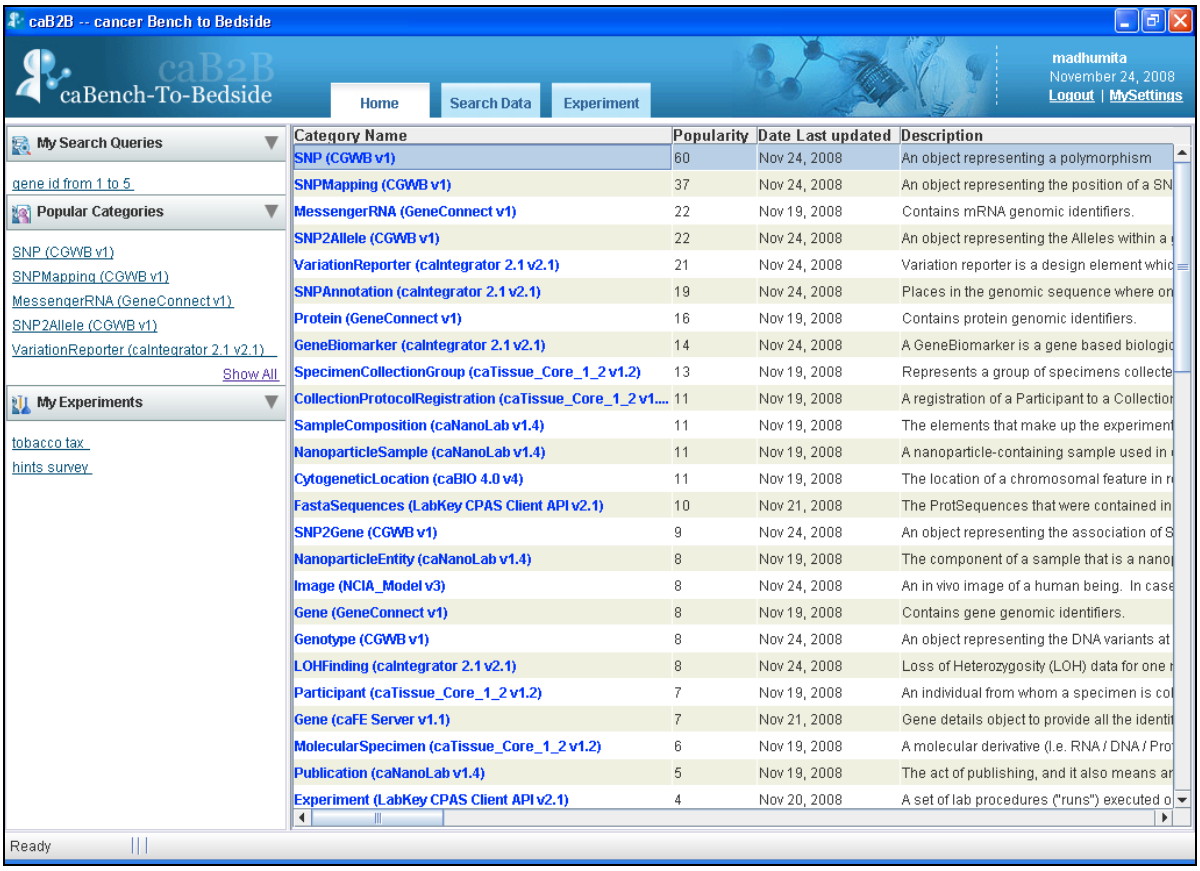


Figure 1.4. Popular categories

Overriding Administrator-defined Services

Administrator initially configures services for applications. These services are available for query in the Client Application. If you want to query services other than the ones configured by administrator, you can do so using *My Settings* link available in the Client Application. Overriding administrator defined services mean you can change the services defined by the administrator to query services of your interest. Follow the steps given below to override administrator defined services:

1. Click on **My Settings** link on the home page.

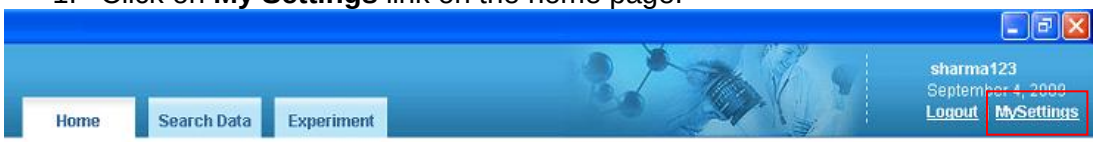


Figure 1.5. My Settings

2. The **My Settings** window displays the various models that have been loaded into the application.

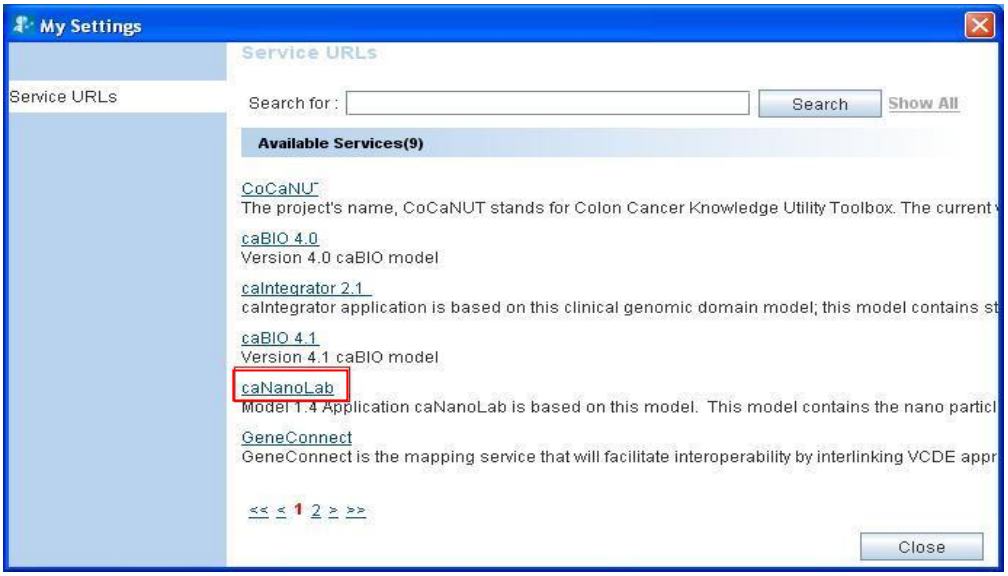


Figure 1.6. My Settings Page

3. Click on a particular model name to display the list of services, example caNanoLab.
4. Select/Deselect the services to override the administrator-defined services.

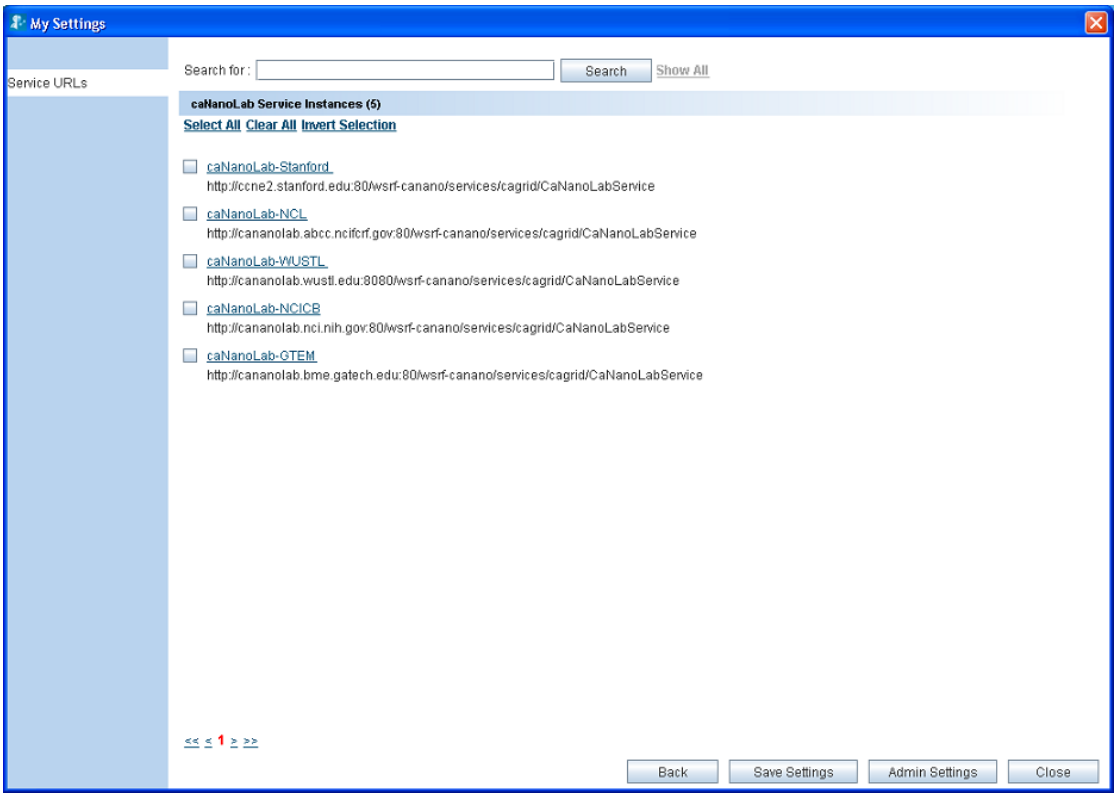


Figure 1.7. Service instances for the selected model

5. Click **Save settings.**
6. To restore administrator defined services click **Admin Settings**.

You might want to retrieve data from a specific instance for a particular query, but not want to apply these settings for all the queries. In such a case, configuring the services during the query is helpful (refer to *Configuring Services for a Query*).

Understanding Categories and Limits

What is a Category?

A category is a collection of *searchable* attributes from different classes of an application/model. A user can use a category to search, view, and add data to experiments. Administrator defines a category by collecting all the useful searchable attributes from different classes into a single category.

For example, administrator can combine attributes from Participant and Specimen classes of caTissuecore application to make a single category. You can use this category, which will have attributes from both the classes.

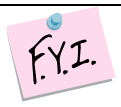
As the category and the class are a collection of searchable attributes, the terms are often used interchangeably in this guide.

The *Gene Annotation* category consists of the following attributes:

- Entrez Gene ID
- Gene Symbol
- Gene Name
- GenBank Accession Number
- Taxonomy ID
- Chromosome

Why do we need categories?

Categories help you to locate all the data elements needed for your search. It helps by combining the attributes from multiple classes.

	Your administrator can create new categories based on your requirements. If there are queries that you perform regularly, discuss them with your administrator.

What is a Multi-model Category (MMC)?

caGrid offers multiple applications to store particular kind of data. You should be able to query the data irrespective of the application used to store it. Multi model category is a solution to address this need. Multi model category is single representation of data that is stored in different applications.
For example, biospecimen data can be stored in different tissue banking repositories like, caTissue Suite or caTissue Core.

Multi-model categories provide you a way to query across such multiple applications Multi-model categories consist of classes from multiple models; these may be different versions of same application or different applications. The attributes from

different models having the similar data correspond to a single attribute in the multi-model category.

	Your administrator will create a multi-model category that contains relevant data elements mapped to each other from respective models. If you want to use a category, which combines multiple applications then, please contact your administrator to define the multi-model categories.

WARNING!	Only the Web Application supports the execution of the multi-model categories. You cannot execute a multi-model category query in the Client Application.

What is a Limit?

A limit is a condition (or criterion or predicate or condition) on a category. It is the smallest unit of a query. For example, *female participants*

What is a Query?

The collection of limits is a query. For example, *female participants who have DNA specimens collected from their breast*. This query has two limits: *female participants, DNA specimens from breast*. There are two categories involved in this query: *Participant and Specimen*.

Chapter 2 Searching Data

This chapter describes how you can use Client Application to fetch data from available caGrid data services like caArray, GeneConnect, caFE, caTissue etc.

The chapter is broadly divided into the following subtopics:

- *Search data – The Query Wizard*
- *Choose Search Category*
- *Define Limit(s)*
- *Define Search Results View*
- *Save Query*
- *View Search Results*
- *Viewing and Saving Data List*

Search data – The Query Wizard

This section explains steps required to obtain data. Click the **Search Data** tab on the **Home** page to display the **Search data for Experiment** wizard. The wizard begins at the Choose Search Category tab. (See *Figure 2.1. Begin search for data)*.

The wizard guides you through the following steps

Step one	Step two	Step three	Step four	Step five
Choose Search Category	Add Limits	Choose an Output Category (Save Query)	View Search Results (Save Query)	View Data List

To search data:

1. Choose Search Category: Search for Category to apply limits.
2. Add Limits: Add limits on the selected Category
3. Choose Output Category (Optional): Select the output format (that is, define the results view). User can also save the query in this step
4. View Search Results: View the results view in a Google-like search results page and add data to the data list (or shopping cart). User can also save the query in this step
5. View Data List: View the data that you have added in the data list and create an Experiment.

Choose Search Category

This step helps you to search the category on which you want to define limits. For example, if you have a query with limits on participants, you can use this feature to find the Participant category.

To find the categories, do the following:

1. Type the search string in the text box (near the top of the screen). Gene, gender, participant, RNA, experiment are examples of search strings.
2. Click **Search** or press ENTER.
3. Search results display the categories that match the entered string as hyperlinks along with a short description. The application to which the category belongs is included in the search result.
4. Once you find a category of your choice, click the desired search result to proceed to the next step.

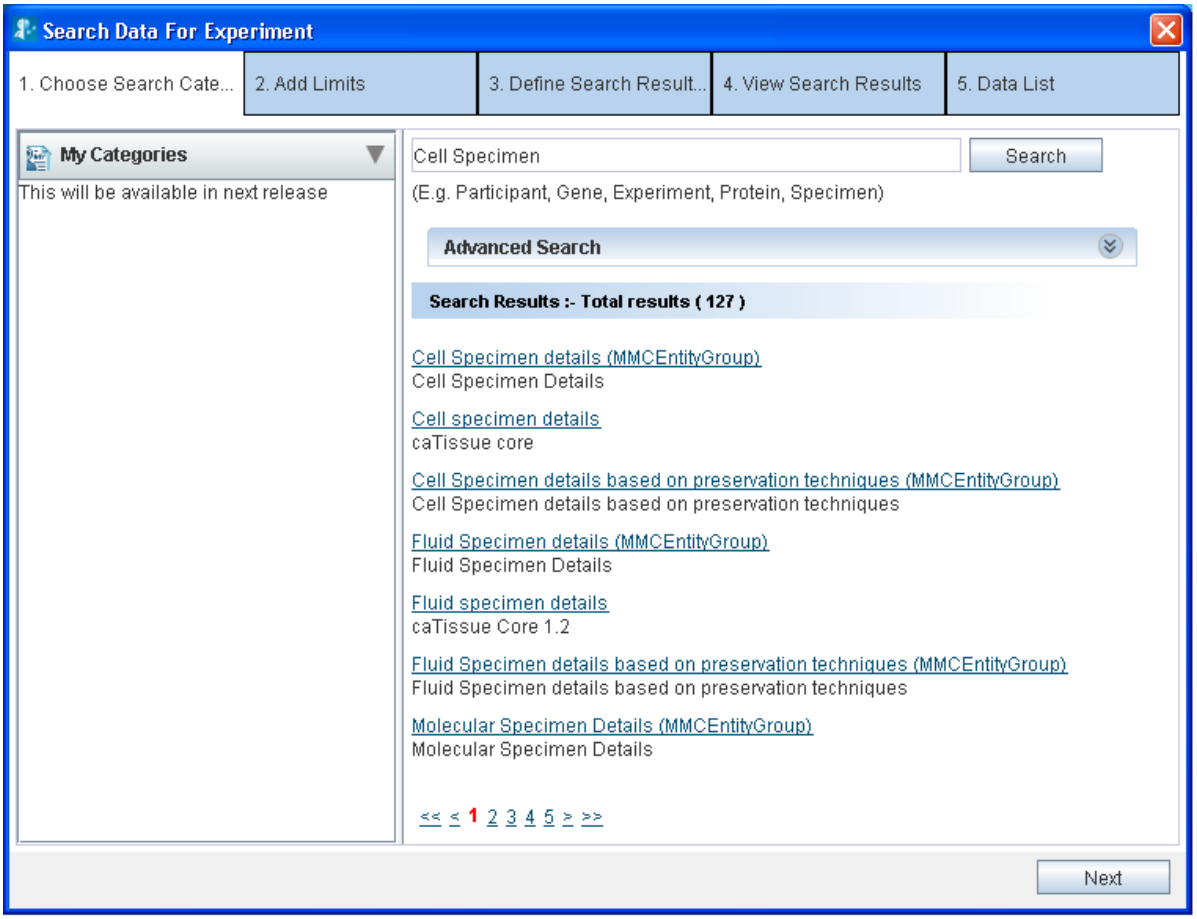


Figure 2.1. Begin search for data

F.Y.I.	Multi-model categories have **MMC EntityGroup** as the application name. In the Client Application, you cannot execute query using the categories that belongs to MMC EntityGroup.

More on searching:

- The search performed is partial and is not case-sensitive. Hence, the search for terms such as *participant, PART or PARTICIPANT* returns the same results.
- If a category description is long, you can view the complete description as a tool-tip by positioning the mouse over that description.
- If the number of matching categories is large, the system displays the results in multiple pages. Click the page numbers at the bottom to navigate through different pages.
- You can separate the multiple search terms by spaces or tabs. The system searches for each term individually and displays a collective set of matching categories. For example, if you enter the search string *Participant Gene*, the result would contain categories that match either *Participant*, *Gene*, or both.

Advanced Search

If your search returns too many results, you may use the Advanced Search. Expanding the **Advanced Search** box displays the advanced options that you can use to refine your search.

The *Figure 2.2* depicts the default selection of the various options

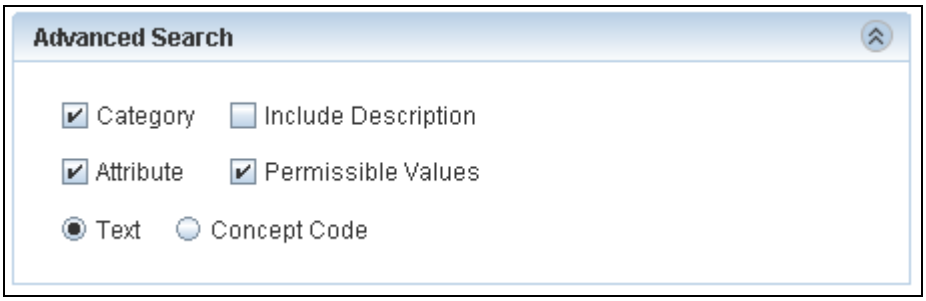


Figure 2.2. Advanced Search

Table 2.1 Advanced Search Options (Checkboxes and radio buttons)

Option	***Description***
Check Boxes	
Category	Includes the category names in the search
Attribute	A category contains many attributes. Select this check box to include the individual attribute names in the search. For example, Select *attribute* and search for *gender*. You will find the category Participant since it contains an attribute by name *Gender*.
Permissible value	Some attributes can have fixed set of values. For example, *Gender* can be *Male, Female or Unspecified. Organ* can be *heart, lung, breast,* and so on. **Permissible Values** is a set of all such values. Select this check box to include the permissible values in the search. For example, if you select the permissible value as *Male* the system displays

	the Participant category since it contains an attribute *Gender* that contains a permissible value *Male gender*.
Include Description	Includes the description text in the search **Note**: **Include Description** check-box is enabled only for a **text** based search when you select the **Category** check box**,** the **Attribute** check box, or both.
Option Buttons	
Text	The system matches the search string that you type with the name of the category, its attributes, or the permissible values based on the text. This is simple, most commonly used and the default way of searching..
Concept Code	The system treats the search string as an **E**nterprise **V**ocabulary **S**ervices (EVS) concept code and matches it with the concept code of the category, its attributes, or the permissible values thereby showing the matching categories. More details on the concept code can be found at http://bioportal.nci.nih.gov

You can make additional refinements with these options as illustrated in the *Appendix A: Examples of category search.*

Define Limit

A limit is a condition (or criterion or predicate) on a category and multiple limits together form a query. Table 2.2 outlines two example queries and the limits added to them.

Table 2.2. Examples of Limit

Example	***Category***	***Limit***
Get the details about all the African male participants	Participant	Gender equals Male Race equals African
Biospecimens from prostate adenocarcinoma	Specimen Characteristics	Tissue Site equals prostate
	Specimen Collection Group	Clinical diagnosis equals adenocarcinoma

Screen for define limit is shown below. It has three parts,

1. Category search is present on the left hand part. It is explained in section *choose search category*
2. *Define Limit* present at right hand side top is explained in section *Steps for defining a limit*
3. *Limit set* present at right hand-side bottom is explained in section *Limit Set – Diagrammatic View*

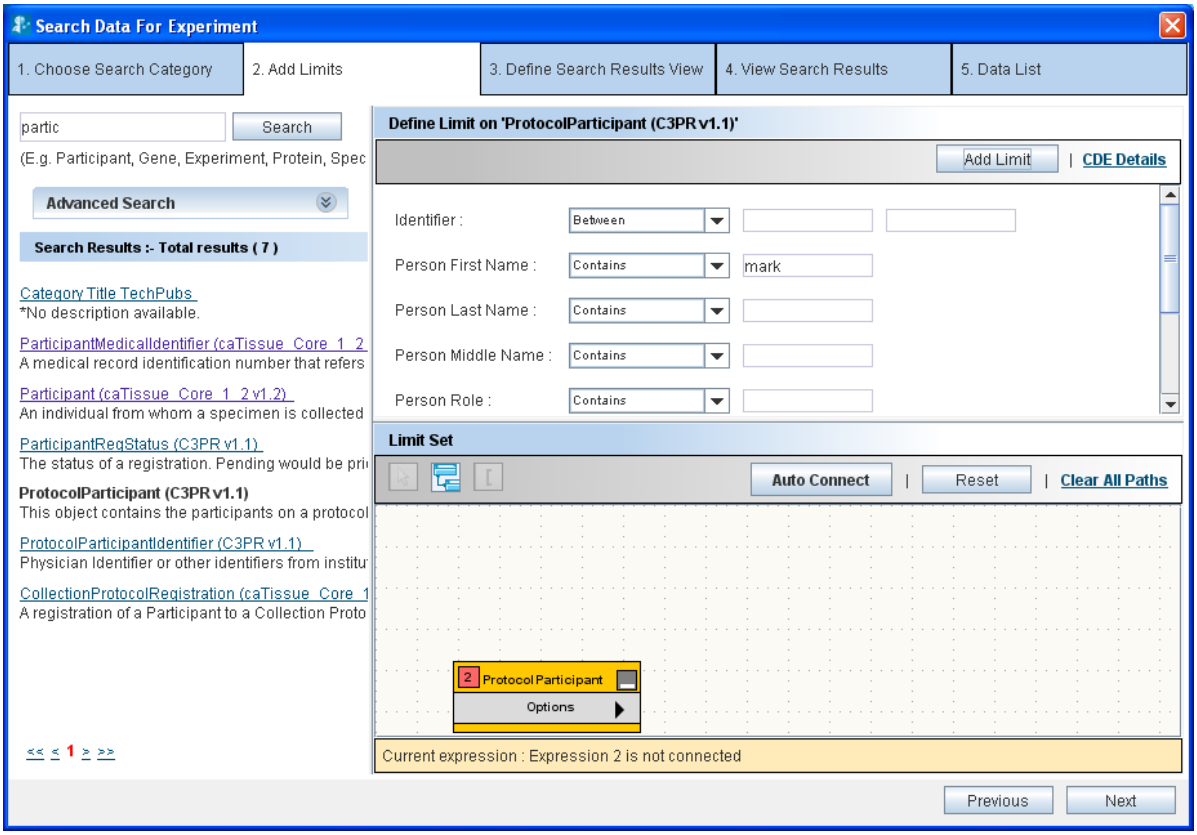


Figure 2.3. Define Limit

Steps for defining a limit

Perform the steps below to define a limit:

1. Choose an operator from the list of operators adjacent to the attribute you wish to constrain.
2. Specify the constraining values.

F.Y.I.	For a detailed explanation of the various operators and their values, refer to *Appendix B: Define Limit: Operators and Values.*

Example:

Single constraint – Show all male participants

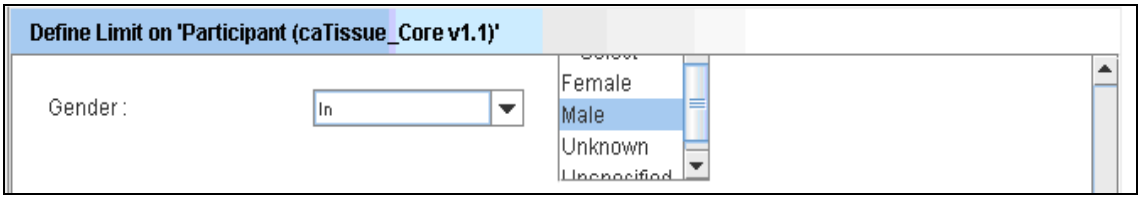


Figure 2.4. Male participants

Example:

Multiple constraints within one limit: Show all male participants who are alive. A user can define multiple conditions on an attribute by pressing CTRL key while selecting different options: Show all specimens, which are either Serum or Plasma.

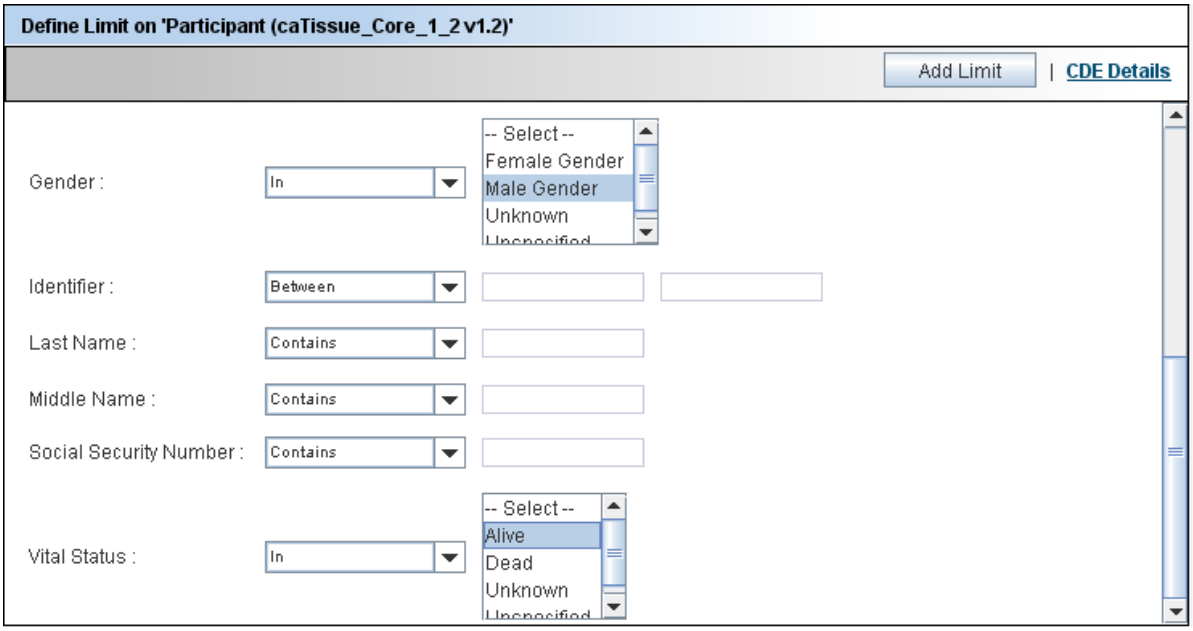


Figure 2.5. Male participants who are alive

WARNING!	Ensure that you click **Add Limit** at the top of the **Define Limits** screen before proceeding. In case you do not click **Add Limit**, the constraints that you have specified are not saved.

Limit Set – Diagrammatic View

Once the limits have been set on the category, clicking on the Add Limit button adds the category in the form of a yellow rectangle (also called a node) into the Limit Set panel. The Limit Set panel consists of the **DiAG**rammatic view (DAG view) panel. DAG view is like a canvas where all the categories on which limits appear in the form of nodes (*Figure 2.6*). On addition of a new limit, the system adds a yellow rectangle to the DAG that represents the limit. As you continue to add limits, the system adds corresponding rectangles to the DAG view.

These rectangles in the DAG view provide a graphical view of all the limits present in the query.

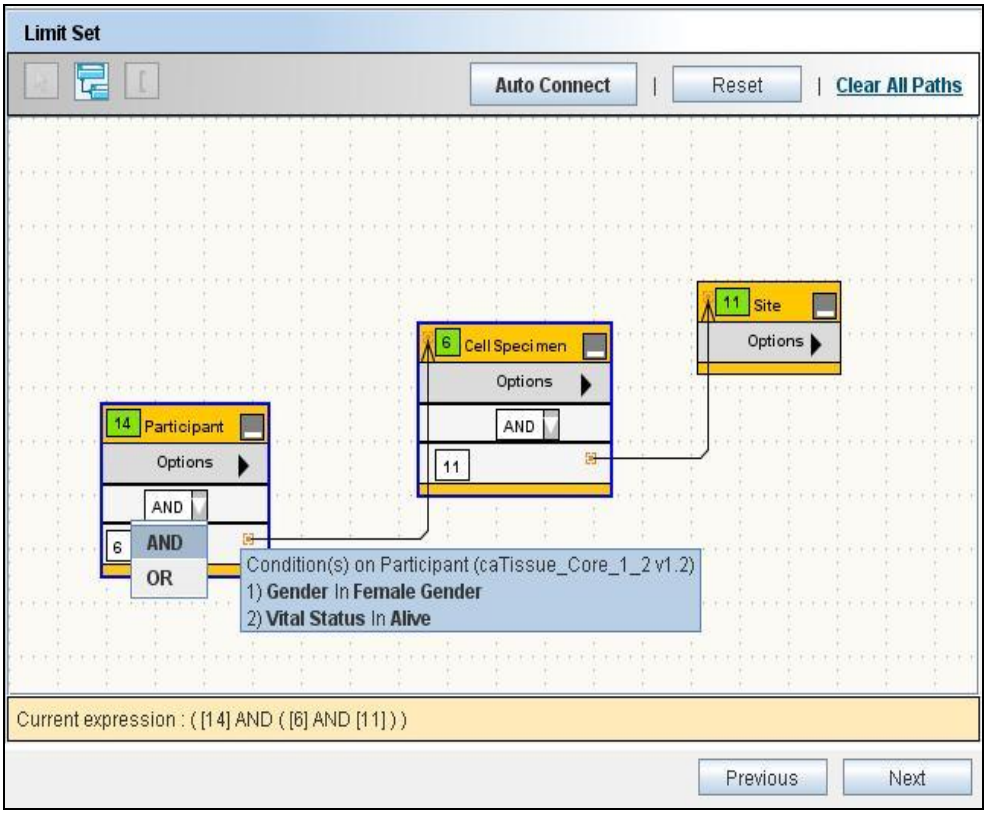
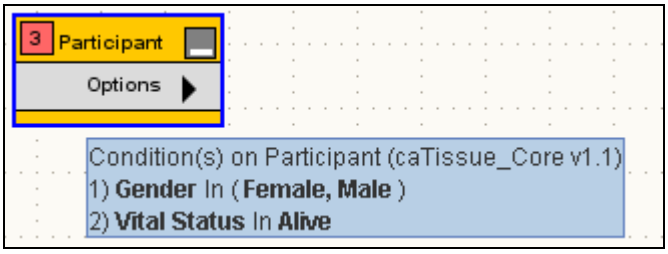


Figure 2.6. Limit Set Panel

DAG Operations

You can perform the following operations within the DAG view.

Table 2.3

View details of the limit	Position your mouse over a node to view the details of that limit. 
Edit the limit	Click Edit Limits from the Options menu of the node whose limits you wish to edit. Alternatively, double click the title of the node. This opens the limit in the Define Limit panel in edit mode. Edit the constraints in the **Define Limit** panel, and click **Edit Limit**.

Delete the limit	Select the nodes that you wish to delete and press DELETE. Alternatively, you can delete a single node by selecting **Delete Limits** from the **Options** menu of the node as shown above.

Connecting the Limits

You must connect all the limits before proceeding to the next step.

When you specify multiple limits, you need to connect these limits to form an expression. Two limits may be connected by using either the **AND** or the **OR** operator. Following is an example of how the DAG looks when the limits are connected.

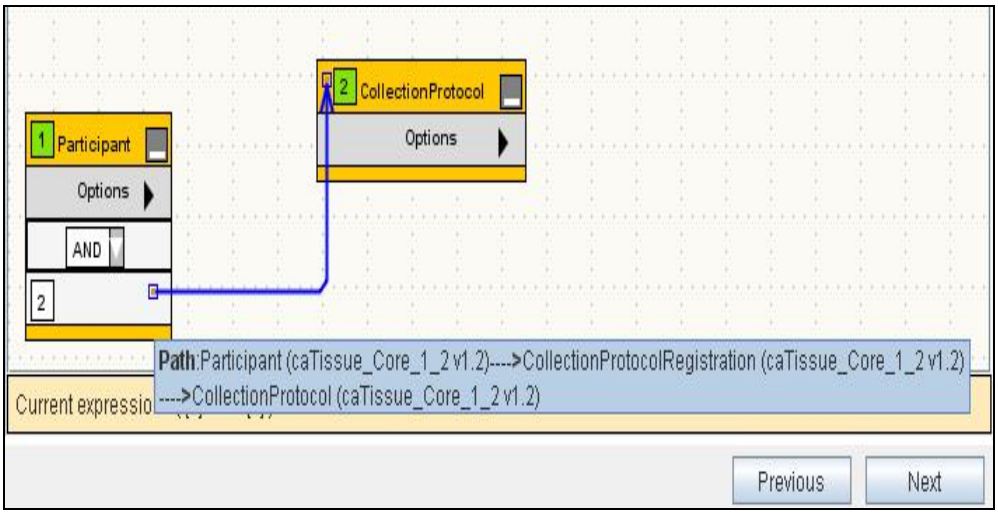


Figure 2.7 Connected Nodes

To connect the limits, you need to specify the following:

- The path to be taken in connecting the nodes
- The logical operator (**AND**, **OR**) in the dropdown box in a node

The rest of this section focuses on various ways in which you can specify the path(s)

What is a Path?

A path tells the system how two categories are to be connected. A path is a way to travel from one category to another. There could be multiple ways to connect two categories and the results of the query differ based on the path chosen.

For example, consider the two categories *Specimen* and *Site* (from the caTissue Core application). A *Specimen* is associated with the *Site* category in two ways: 1) Site where the specimen was collected 2) Site where the specimen is stored.

That means there are two paths between *Specimen* and *Site*: 1) Through Specimen Collection Group for the collection site.2) Through Storage Container for the storage site.

Query example 1: Show all the available specimens collected in Barnes Jewish Hospital. In this query, as you are specifically interested in a particular collection site, you should choose the path through Specimen Collection Group.

Query example 2: Show all the available specimens stored in Barnes Jewish Hospital. In this query, as you are specifically interested in a particular storage site, you must choose the path through Storage Container.

Therefore, it is important to choose the right path for your query.

Connecting Limits

Click **Auto Connect**. The system automatically connects all the boxes based on the settings provided by your administrator.

However, the administrator might not have configured the path between all the categories. In such cases, you can manually connect the limits as described below.

Example query: For all the male participants, show all the available specimens collected in Barnes Jewish Hospital.

1. After adding three limits on the *Participant*, *Specimen* and *Site* the DAG view looks as shown in the figure below:

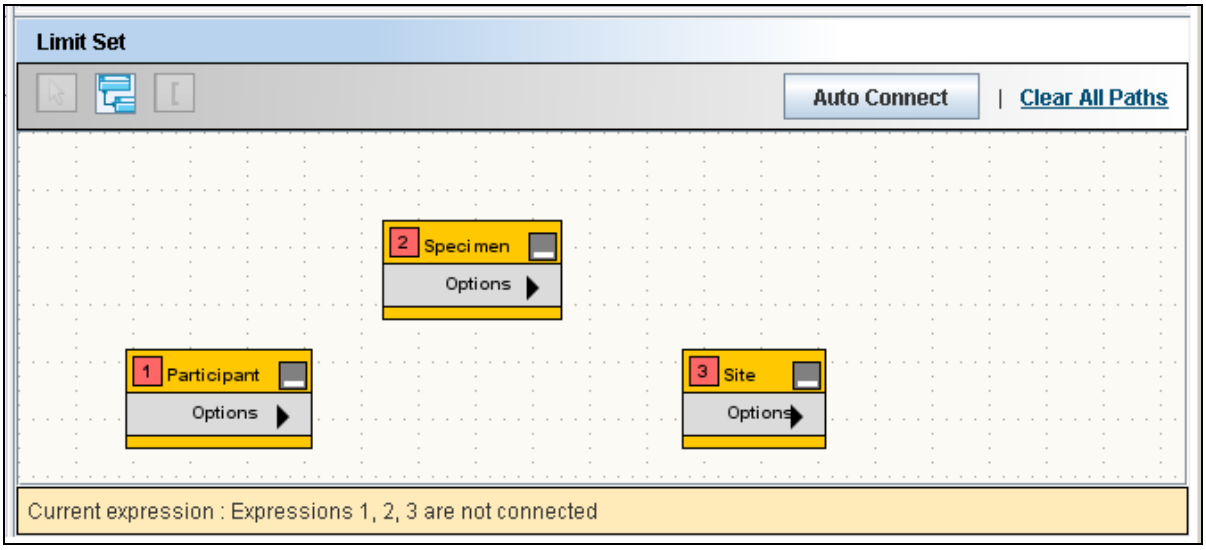


Figure 2.8. Limit Set

2. Click the *Participant* and the *Specimen* rectangles to select them.
3. Click **Connect Limits** to connect the two nodes. The resulting DAG view looks like *Figure 2.9*.

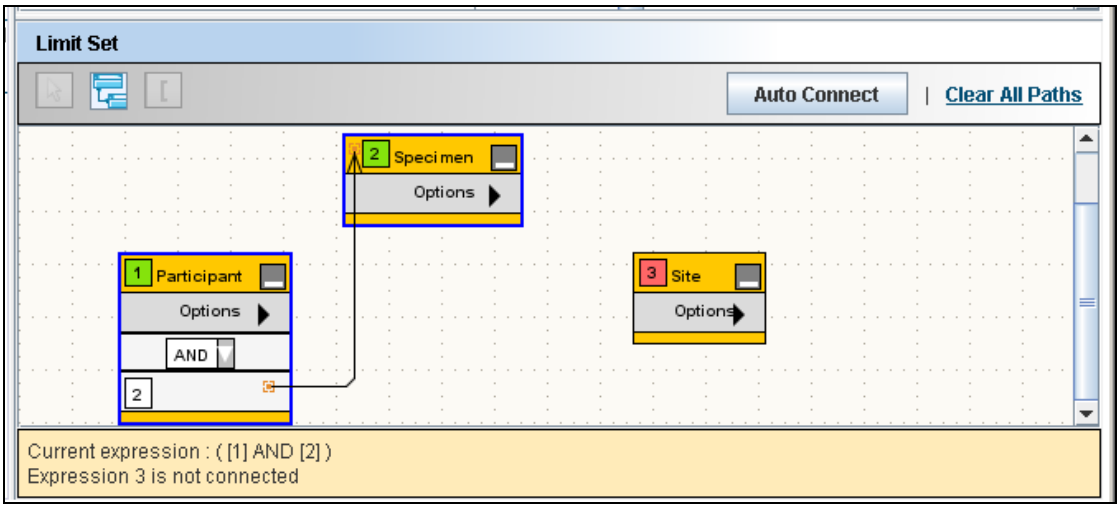


Figure 2.9. Limit Set with Two Connected Limits

Note: If there is only one path between *Participant* and *Specimen*, the system will automatically connect it. Default operator is AND, it can be changed to OR using operator dropdown.

4. Select the *Specimen* and *Site* rectangles and click **Connect Limits**.
5. Since there are more than one path between *Specimen* and *Site*, the system cannot connect the limits on its own. Therefore, it displays the **Ambiguity Resolver** window as shown in *Figure 2.10. Path Ambiguity Resolver*

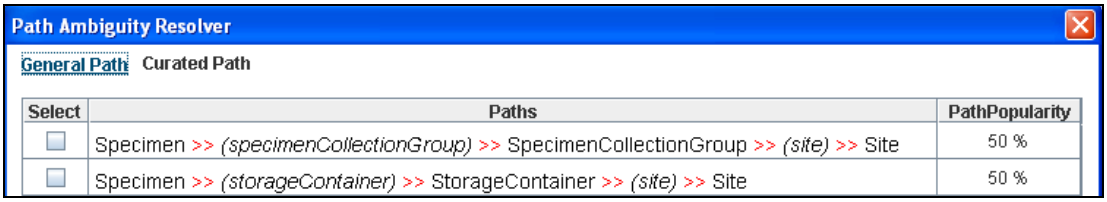


Figure 2.10. Path Ambiguity Resolver

6. Select the first path to find out the collection site.
7. Once the path(s) is determined, the system creates a connection between the two nodes (see figure below). You can position the mouse over the connection to view the path(s) that you chose.

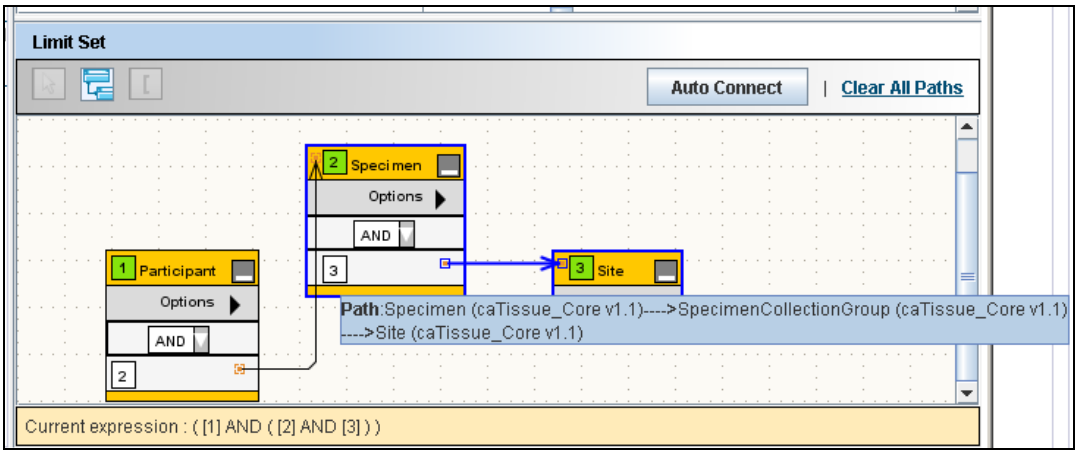


Figure 2.11. Limit Set with Three Connected Limits

8. Specify the operator (**AND** or **OR**) using the dropdown box in the node. For above query the operator will be **AND**
9. Click **Next**.

Deleting a Connection

You can delete a connection by clicking the connection arrow between the two nodes and pressing the DELETE key.

The Information Panel

The **Information** panel is located at the bottom of the **Limit Set** panel.

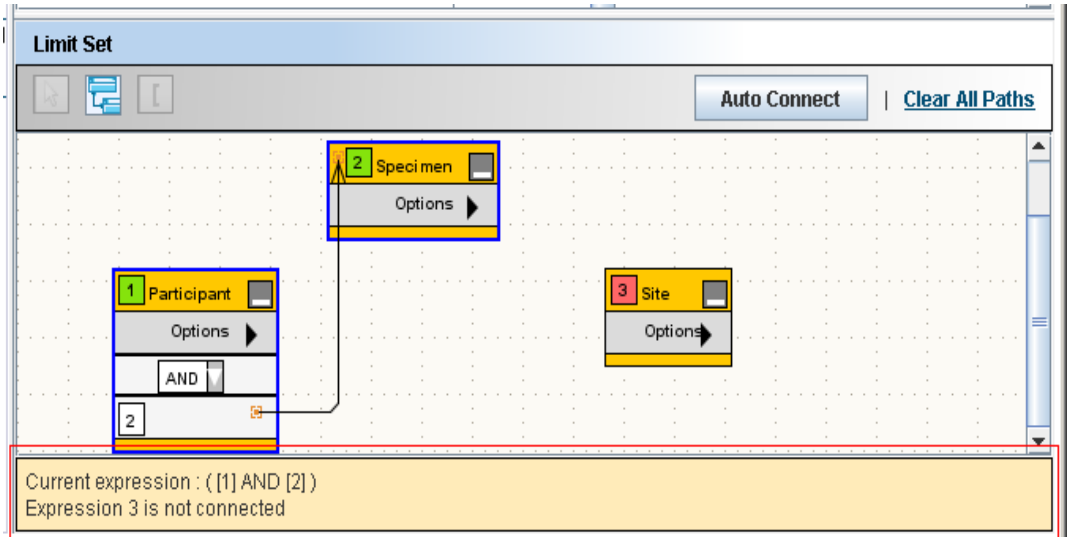


Figure 2.12. The Information Panel

The **Information** panel provides the following information:

- The expression formed because of how the limits are connected.
- Limits that are not yet connected to any other limit

- To execute the query, there should not be any unconnected rectangles.

Reset, Previous and Next Buttons

These buttons are located on the Limit Set panel (refer to, *Figure 2.6. Limit Set Panel*)

- The **Reset** button clears all the specified limits.
- The **Previous** button takes you to **Choose Search Category** step of the wizard.
- The **Next** button takes you to the **Choose Output Category** step of the wizard. A valid query allows you to proceed to the next step. System displays an appropriate error message if the query is invalid.

Define Search Results View

After specifying the limits, you proceed to choose the output category for the data. For example, “show all the participants who had biospecimens collected from their breast”. *Participant* is the category that you want to view, based on a limit on the specimens.

The **Define Search Results View** tab allows you to select the output category for the query.

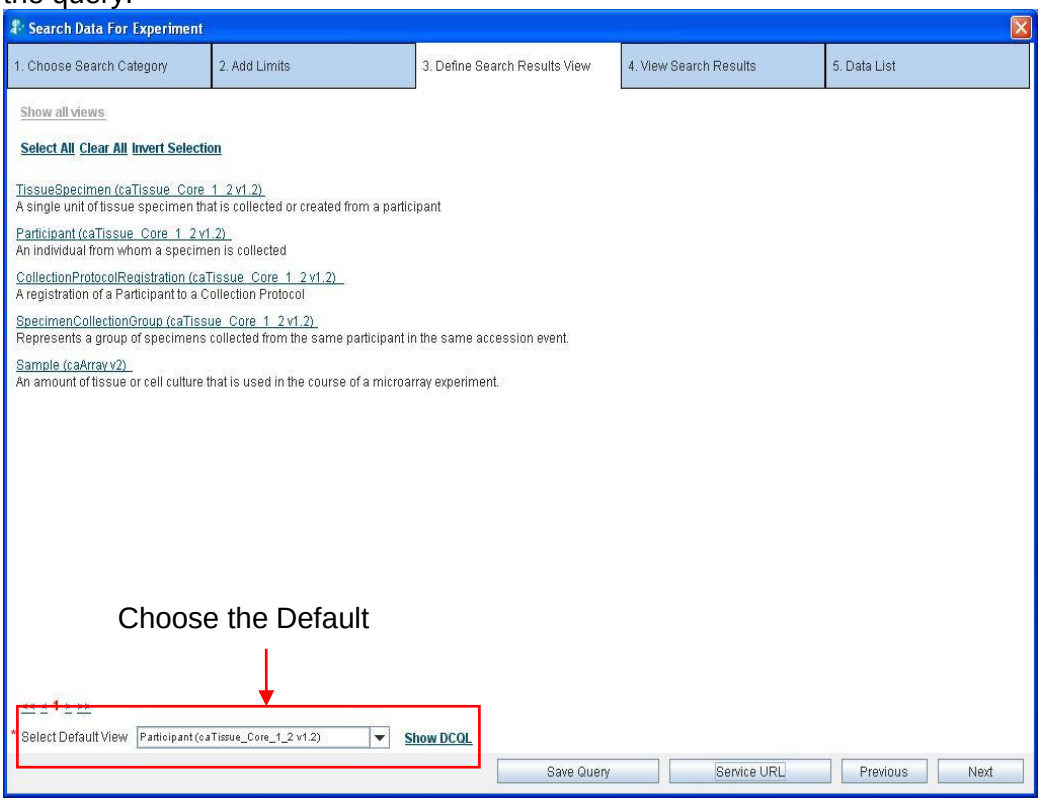


Figure 2.13. Define Search Result View.

As shown in *Figure 2.13*, the **Select Default View** dropdown box at the bottom of the screen shows all the categories present in the query. From the **Select Default View** dropdown box, select the category whose data you wish to obtain.

The Next, Previous, Save Query and Service URL Buttons

- To edit the limits, click **Previous** and go to the Define Limit step
- Click **Next** to execute the query that you have created
- Click **Save Query** (refer to *Save Query*) to save the query.
- Click **Service URL** to modify the services.

WARNING!	On clicking the Next button, the system executes the query and fetches the data from appropriate data services over the caGrid. This step could be time-consuming. It may take few minutes to couple of hours based on the network speed and complexity of the query.

Configuring Services for a Query

A query fetches data from all the services configured by you using My Settings. If you have not selected the services, then services selected by the administrator will be queried. For a particular query, you might want to get information from specific instances and for some other query; you might need some different instance. For such cases, Client Application allows you to modify the services for a particular query just before executing it.

To select the services while executing a query follow the steps given below:

1. Click **Service URL** button on the **Define Search Results View** page.

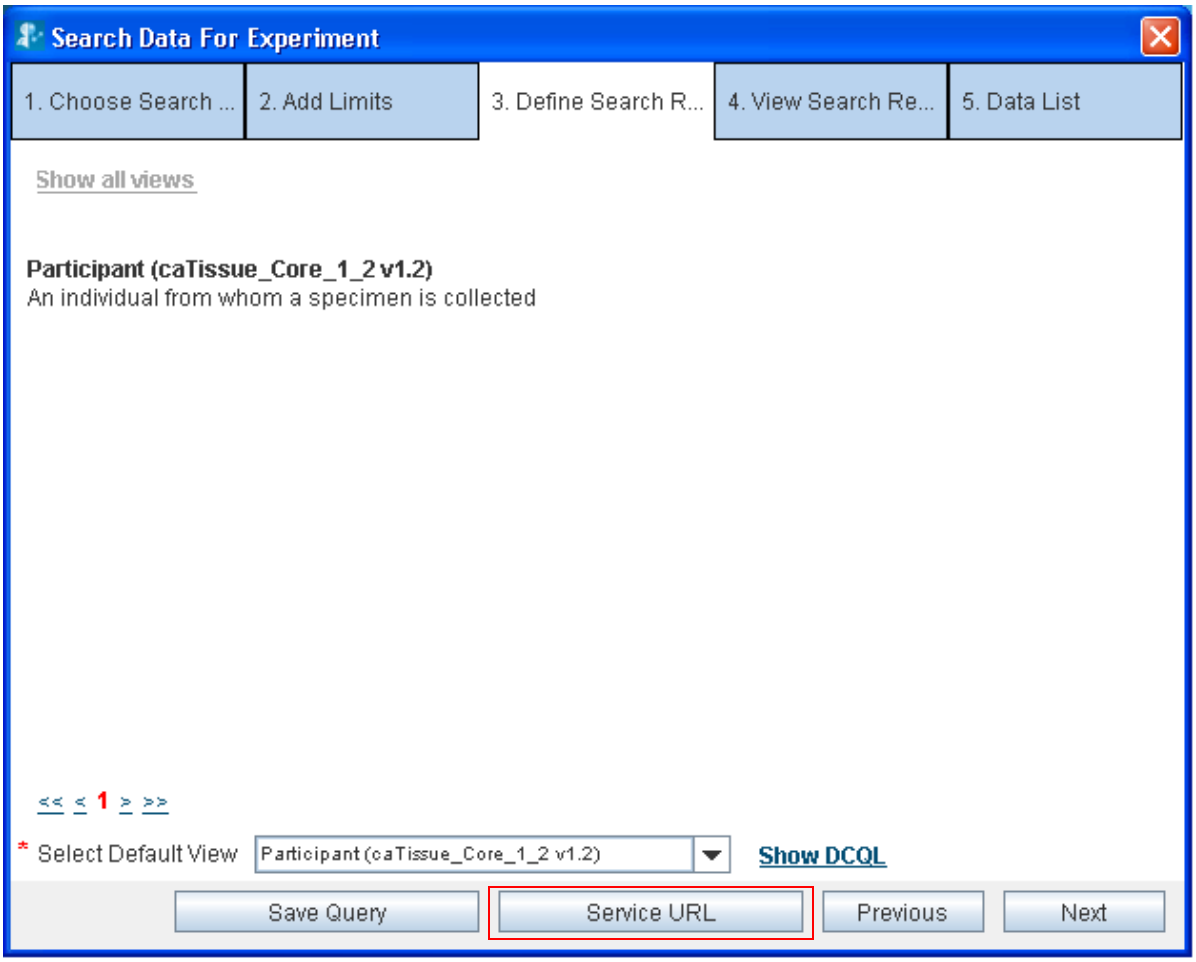


Figure 2.14. Service URL button on the query page.

2. Click the model name.

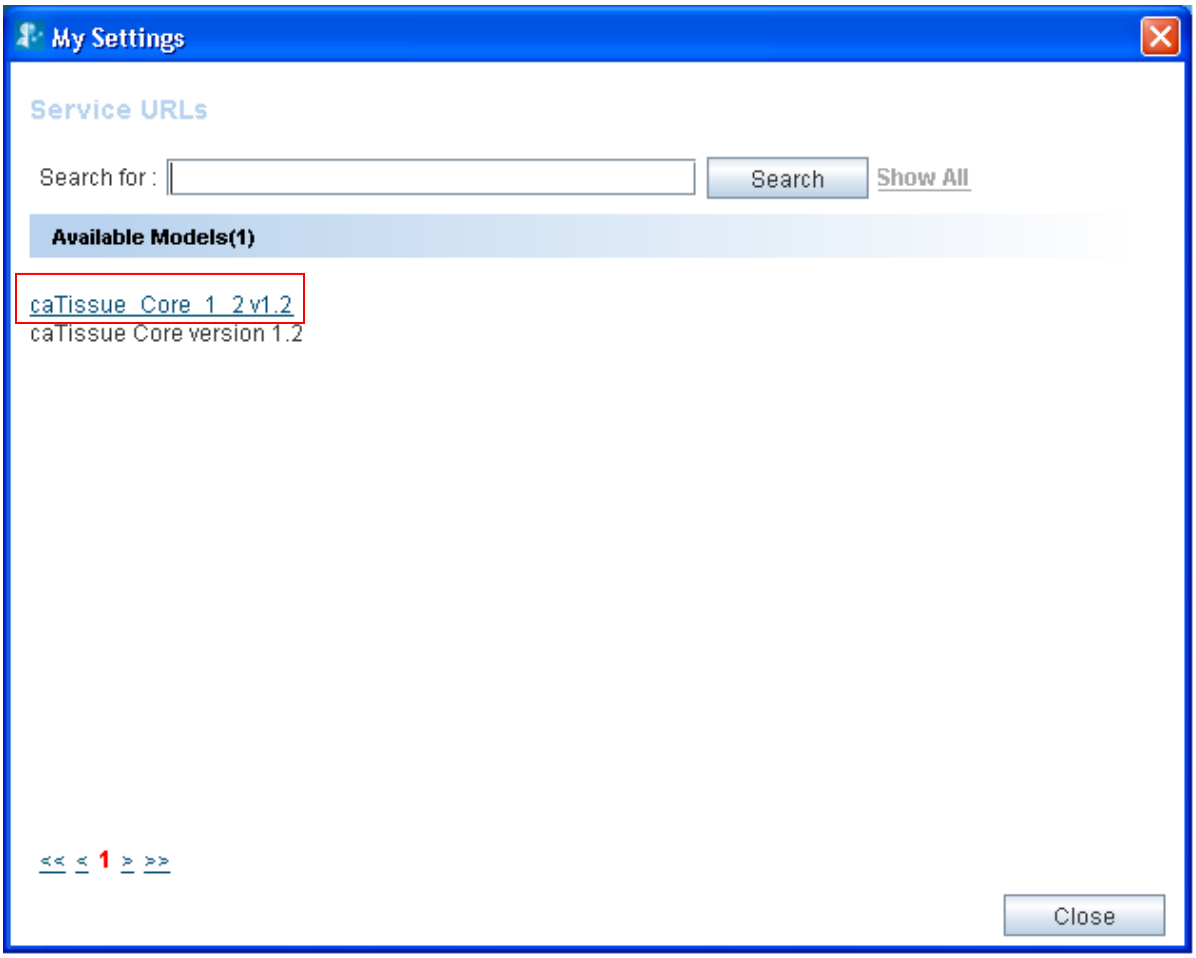


Figure 2.15. Showing all the models available in the query

3. Select the services.

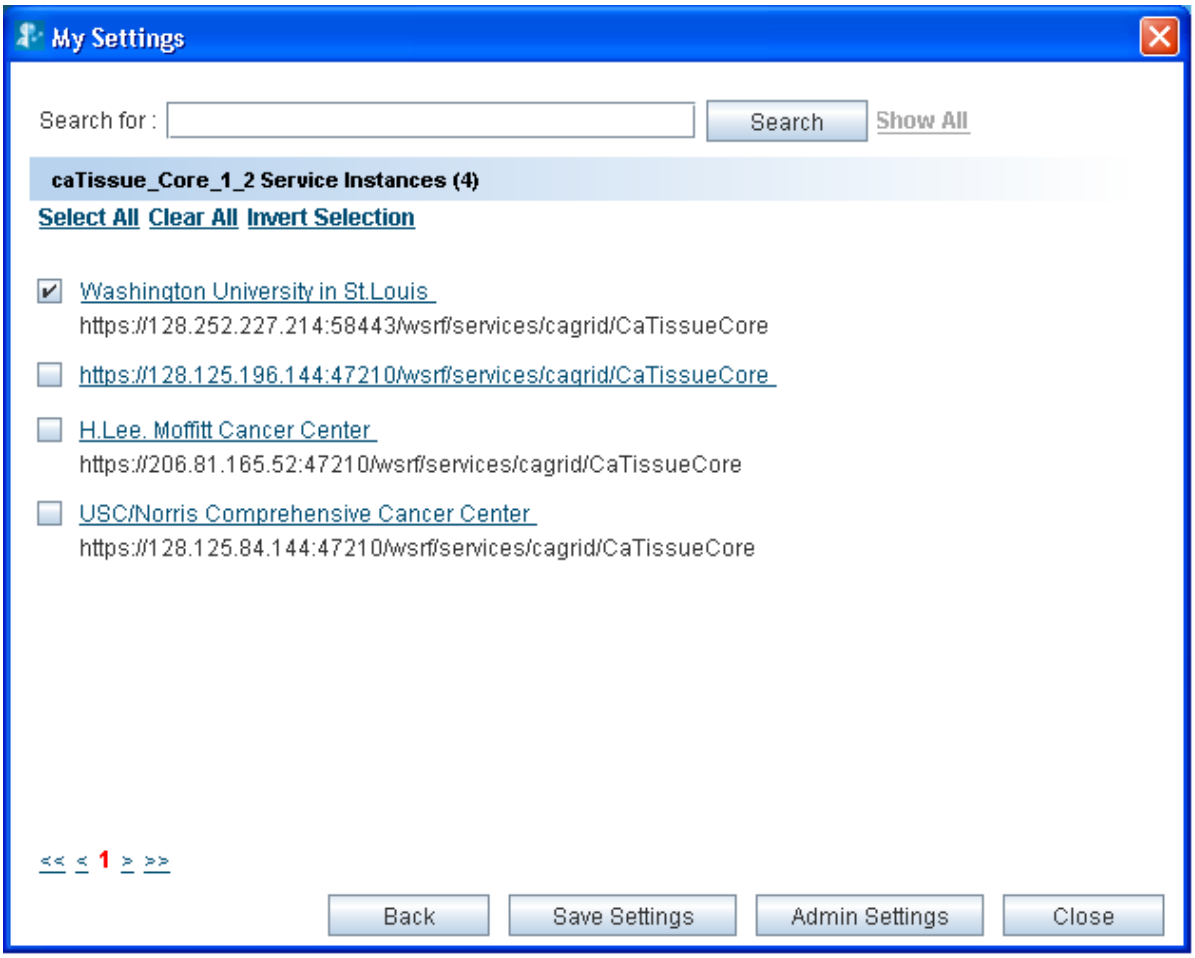


Figure 2.16. Service URL selection

4. Click **Save.**

If your query involves categories from more than one model, then you can set the services for all the models involved in your query.

	The services defined at this step are applicable to this query only. This does not modify the services you configured using My Settings. These settings are not applicable for the next query.

DCQL Viewer

DCQL is the query language used by caGrid data services to express queries against a data source using an object-oriented language. It is the query language used by the caGrid to retrieve data from various data services. It is in the form of XML schemas. A user who is new to caGrid can view the DCQL and learn it to use the language.

To view the DCQL on the Define Search Results view tab click on the **Show DCQL** link. The pop-up window shows the DCQL executed for the query.

You can click **Export** to download the DCQL query in the XML or TXT format.

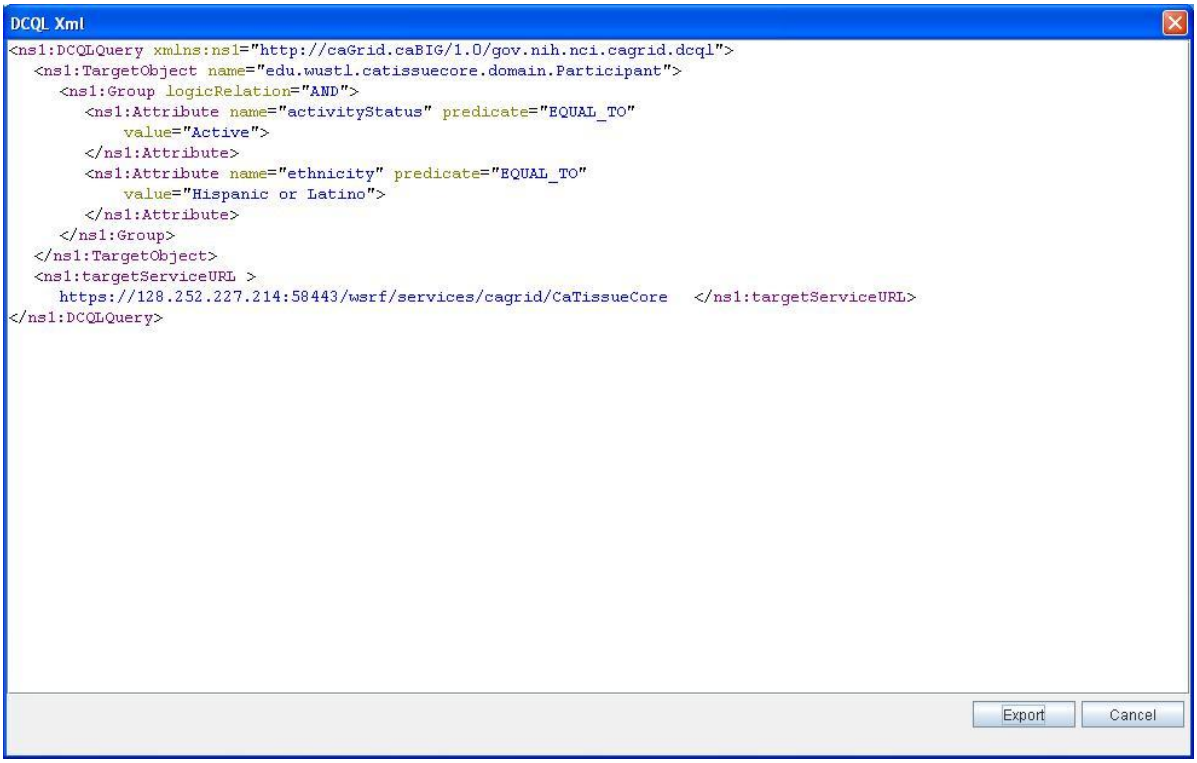


Figure 2.17 DCQL query

Save Query

This option helps you to save the query created using the DAG. While saving a query all the classes present in the query get saved along with the conditions defined on them.

Why to Save Query?

A saved query helps you to view the result of the query without having to create the query in DAG. You can create the query in DAG once and save it for later use. You can parameterize a saved query by saving conditions on the attributes present in the query (described further below). You can execute a saved query with or without changing the conditions defined in the saved query.

Advantages of Saving a Query:

- Saving complex queries with multiple nodes
- You can execute the saved queries later to get the results. Saved queries created by one user will not be available to another user.
- Setting up parameterized queries that allow you to use the same query for a variety of conditions on a single attribute

- Administrator for the *Web Application* can save query as a keyword query or saved search for *Web Application* users.

Types of Saved Queries:

There are two types of saved queries available in the Client Application:

- Non-parameterized queries: These queries do not have any parameters defined. It means that you cannot modify these queries at the time of execution.
- Parameterized queries: These queries have parameters defined while saving it. You can modify the parameters at the time of query execution to generate a new query.

Parameterized queries for caB2B Web Application user:

The queries saved through Client Application are available in the *Web Application*. There are two types of queries available in the Web Application:

- Saved Search: A saved search is a normal parameterized query saved in the Client Application. You can execute it by defining conditions on the available parameters. Application saves the queries along with the operators you selected. Hence, if you defined AND operator between a pair classes and an OR operator between another pair, the query will have the AND and OR operator between that set of classes. The **Saved search** section displays all the available parameterized queries.
- Keyword Search: A keyword search is a parameterized query with OR operator defined between the attributes and the classes irrespective of the operator defined by you. In addition, a keyword query saves only the string attributes from the list of user-defined parameters.

	The Saved search section of the Web Application displays the available parameterized query created by an administrator. The keyword queries are not visible in Web Application. They are executed when you enter a keyword in web application (refer to *End User Manual of Web Application*)

How to Save Query?

To save a query, add limits on categories and connect them via the available paths. After forming the query navigate to the Define Search Results View tab, which provides you the option to save a query. Use the **Save Query** option at the bottom of this tab to save the query. The Save Query option is also available on the View Search Results tab.

How to save a non-parameterized query?

1. Click **Save Query** (as shown in *Figure 2.13. Define Search Result View.*).
2. Enter the **Title** and **Description** for the query to be saved, in the save query panel (see the figure below).
3. Select the **Keyword search, Saved search** or **Both** option based on the type of query to be saved.

4. Click **Save**.

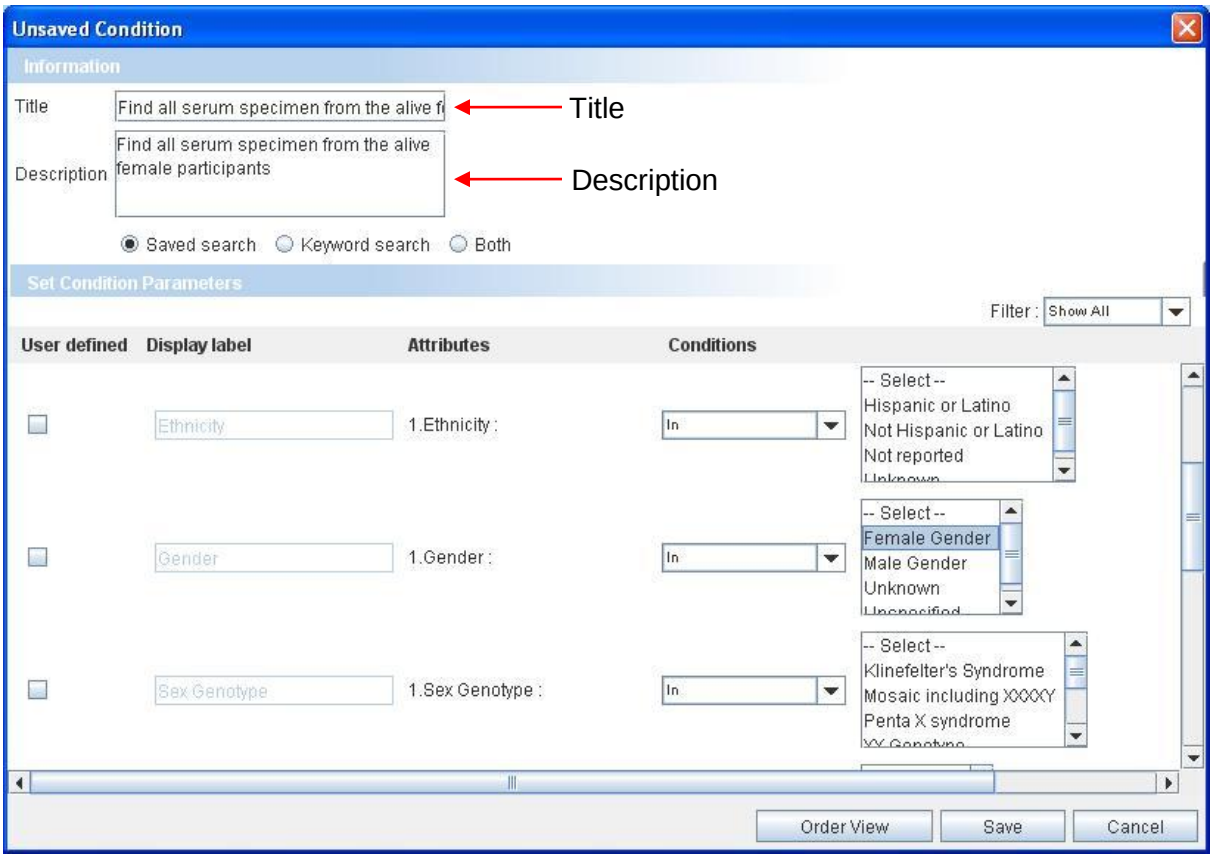


Figure 2.18. Save Query Panel-Non Parameterized Query

How to save a parameterized query?

1. Click **Save Query** (as shown in *Figure 2.13. Define Search Result View.*).
2. Enter the **Title** and **Description** for the query to be saved, in the save query panel (see the figure below).
3. Select the **Keyword search, Saved search** or **Both** option based on the type of query to be saved.
4. Select the **User Defined Check Box** for the attribute you wish to save as a parameter.
5. Use **Filter** option to define condition on any of the attribute of the category. It includes attributes other than those initially defined in the query.
6. Click **Save**.

F.Y.I.	Click **Cancel** to exit without saving the query. If you are the Administrator of the Web Application, your saved query would be visible in the Saved Search panel of Web Application.

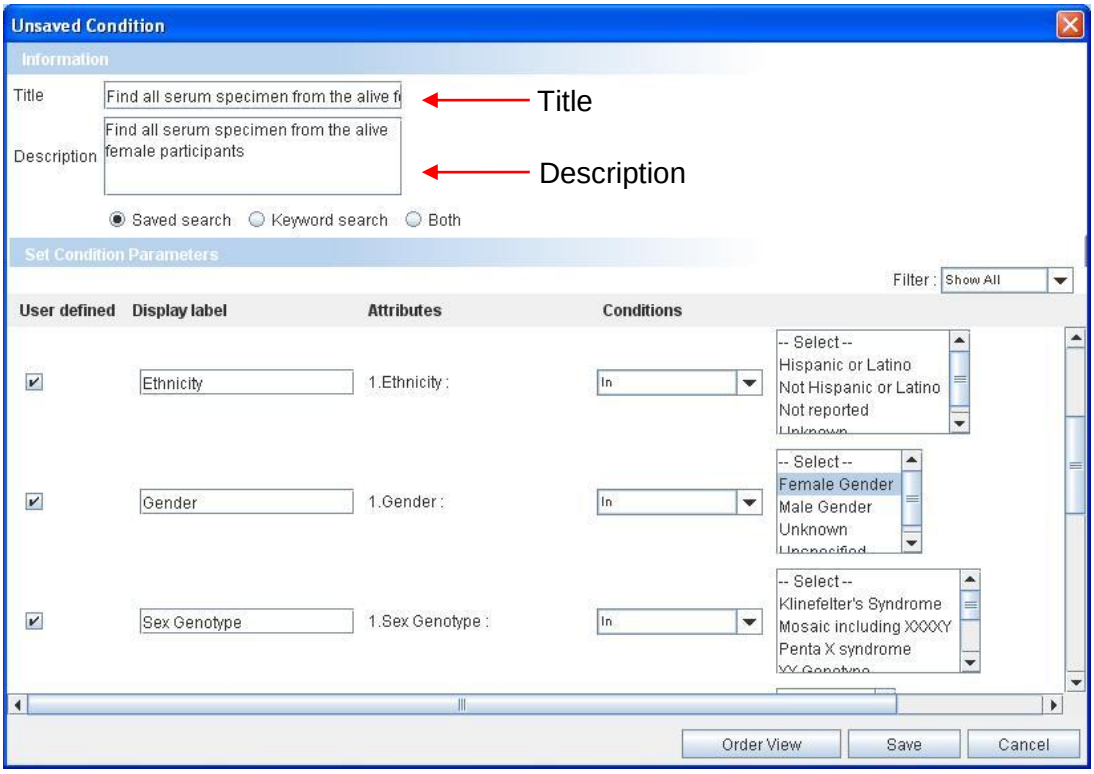


Figure 2.19. Save Query Panel-Parameterized Query

F.Y.I.	The Saved queried with a # in the query title are keyword queries. You can see these queries in the Client Application.

Executing a Saved Query

All the saved queries are visible in the **My Search Queries** stack box in the left hand side panel on the home page. To execute a saved query, follow the steps given below:

1. Click on the title of the query to be executed

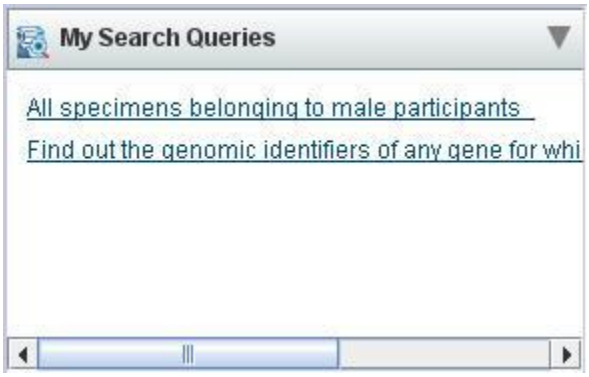


Figure 2.20. Saved Query Stack box

The **Saved Query Panel** opens.

2. For non-parameterized queries, click **Show Results** on the execute query panel

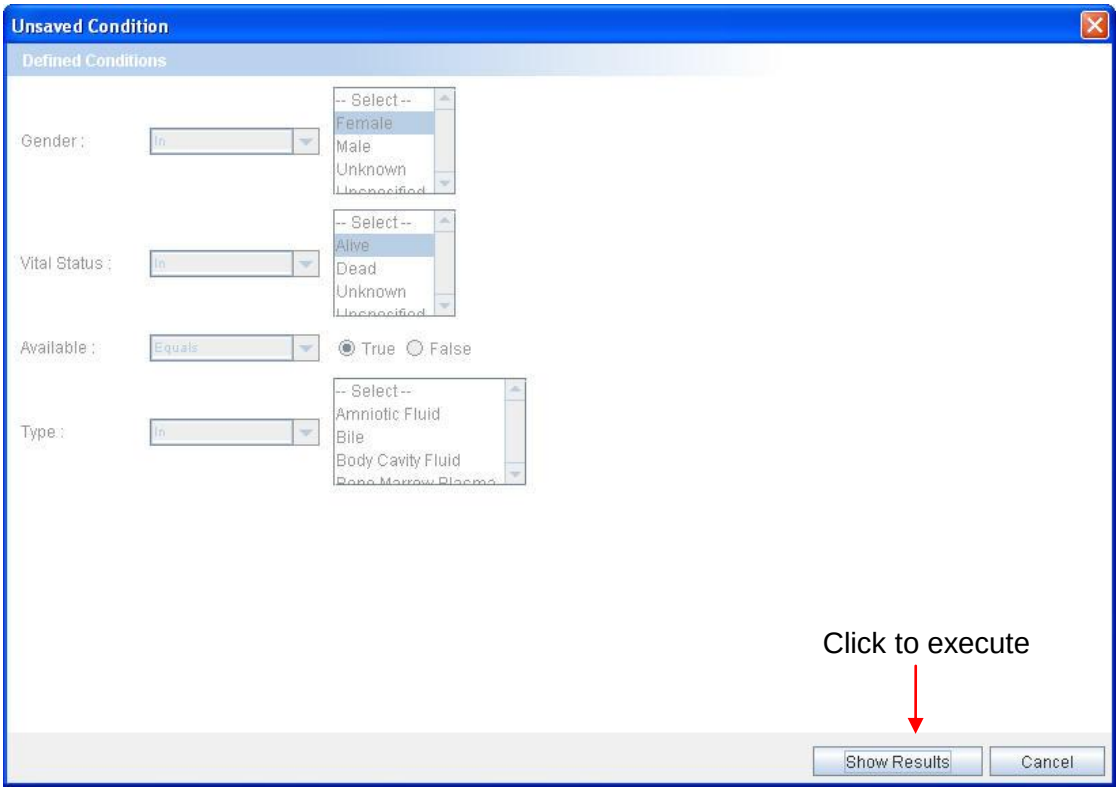


Figure 2.21. Execute Query Panel- Non Parameterized Query

3. For parameterized queries,
   a. Change the conditions for the parameterized attributes
   b. Click **Show Results.**

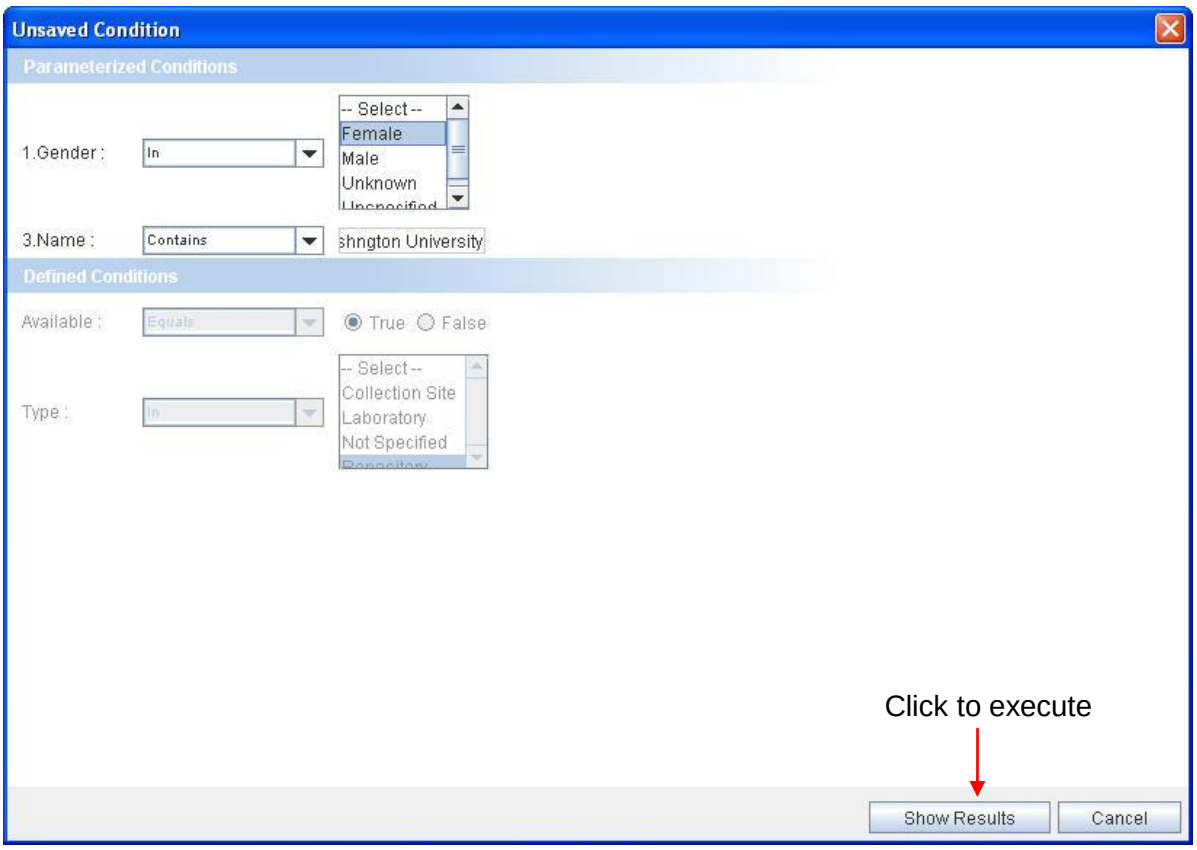


Figure 2.22. Execute Query Panel-Parameterized Query

The save query option provides an easy alternative way of executing the queries without actually forming the query again. The parameterized query allows changing the conditions on attributes thereby allowing flexibility to the saved queries.

Use the Web Application to execute the saved keyword queries (refer to caB2B Web Application User Manual).

View Search Results

The **View Search Results** tab displays the results obtained by executing the query defined in the previous steps. The concept of viewing results in caB2B is similar to viewing results after a Google search. The following are the similarities:

- Client Application returns a set of records that match your limits.
- You can see some details of the record along with the record.
- Click on one of the records to view more details.
- Records provide additional information, for example, from Participants you can see the Collection Protocol in which they are registered. This is similar to Google where, after you click one of the hyperlinks, you can traverse into other web pages by clicking the hyperlinks within that page.

- As in Google, you will see the results divided into multiple pages for easy navigation.

If you find any data of interest, you can add it to the **Data List**.

The following image displays a typical results view. The **View Search Results** tab displays the records that match the limits and the data that you have already added to the data list.

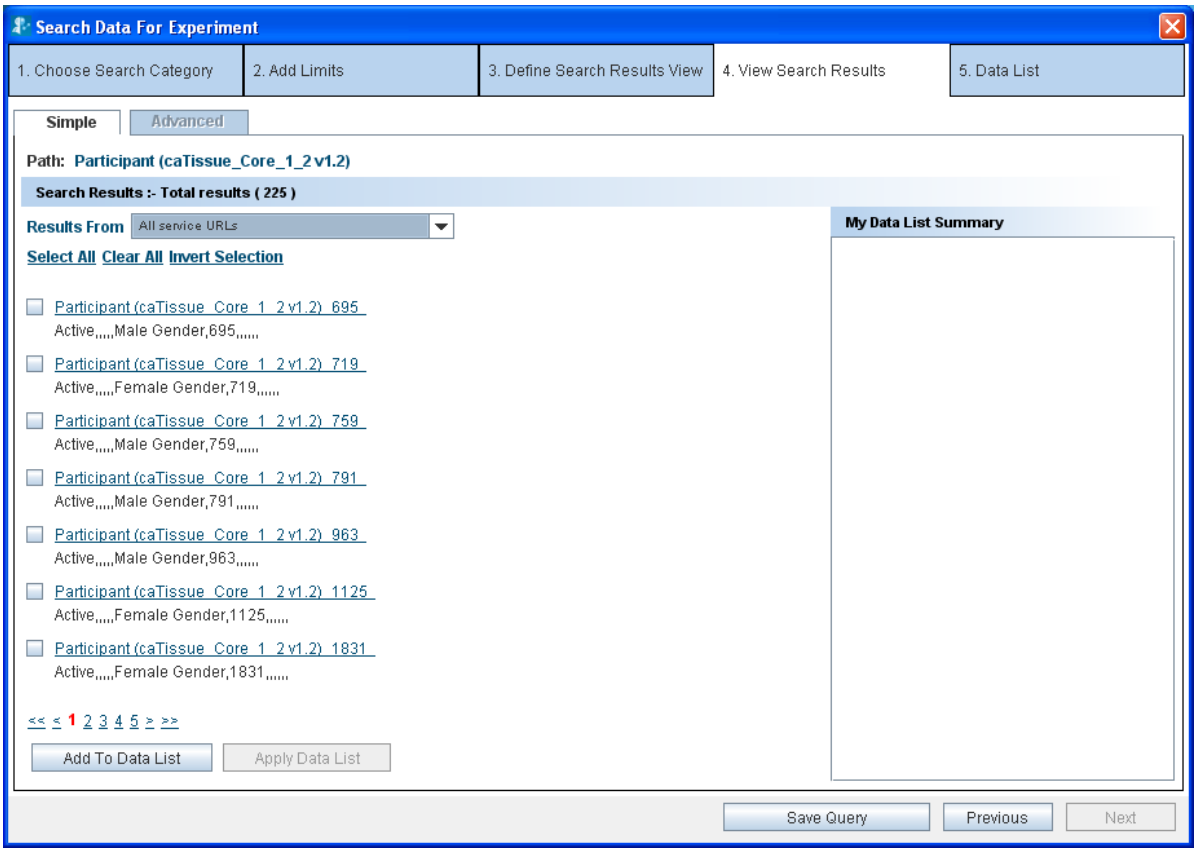


Figure 2.23. View Search Results

	You cannot execute a query on a multi-model category in the Client Application. Application shows you a message stating that: Currently there is no support for Multimodel category query execution in caB2B client .Please use Web Application for executing Multimodel category query

Record Details

Click on the hyperlink of a record on the **View Search Results** to take you to the **Record Details** page. Here you can view the values of all the attributes of the category.

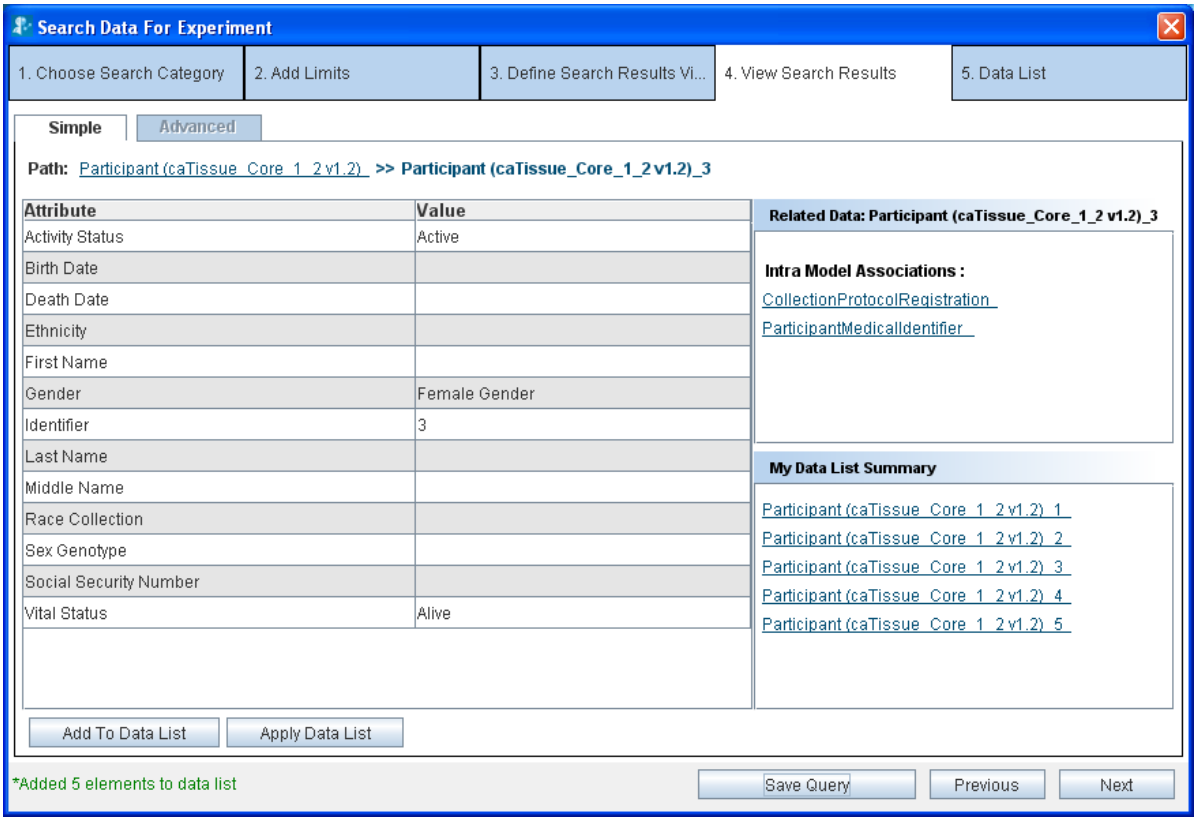


Figure 2.24. Results' Details

Figure 2.14 shows a view of the Results Details page it shows the following sections:

- Value of all the attributes in the record
- Links to related data
- Summary of the records that you have already added to the data list

Related Data

As shown in *Figure 2.24. Results' Details* above, the **Related Data** box on the right displays the categories related to the current category. Click on any of the categories to display records for that category. For example, if you are currently viewing records of Participant (caTissue), the related data box shows the following categories:

- *Collection Protocol Registration (caTissue)*
- *Participant Medical Identifier(caTissue)*

If you now click *Collection Protocol Registration*, you would see the Collection Protocol Registrations related to the Participant you are currently viewing.

Breadcrumbs

Breadcrumbs or a breadcrumb trail is a navigation technique used in user interfaces. Its purpose is to give users a way to keep track of their location within programs or documents. (Definition from Wikipedia entry)

As you traverse the results view, the top of the page displays the breadcrumbs. For example, if you were viewing a molecular specimen record related to a participant, you would see the path as

Participant >> participant_1125 >> collection protocol registration >> specimen collection group >> molecular specimen

This means that you originally queried for participant. Then you fetched related molecular specimen for participant_1125 via collection protocol registration and specimen collection group.

Click any of the links in the breadcrumbs to navigate back to the corresponding page. For example, you can click Participant in the above path to go back to the page with the list of participants.

Adding data to the Data List

A data list is like a shopping cart, which enables you to add interesting data to the shopping cart as you traverse through the results. You can add data to the data list in two ways:

- Click **Add to Data List**
- Click **Apply Data List**

	As you add the data to the data list, the **My Data List Summary** box on the right hand side reflects the updated data list.

Add to Data List

Select the record that you want to add to the data list and click **Add to Data List**. You can either add one record at a time or add multiple records by selecting the checkboxes placed along with the record summary.

Using Apply Data List

Once you obtain the related data for a particular record, **Apply Data List** allows you to fetch similar related data for several other records easily. We outline an example of this approach below.

A typical user query might do the following:

1. Fetch the data for the *Participant* category.
2. Fetch related data from the category *Molecular Specimen* for a record of *Participant,* for example, *participant_1125*.
3. Add the *Molecular Specimen* record to the data list.

Figure below shows the resulting data list.

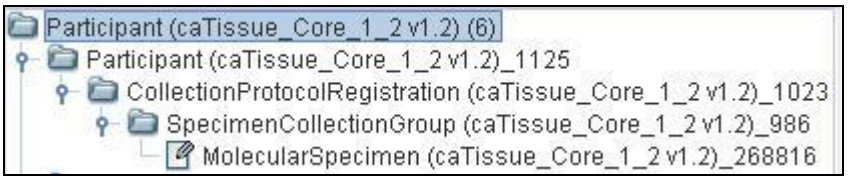


Figure 2.25. Data List

If you wish to obtain the *Specimen data* for other *Participant* records, you have to fetch related *Specimen* for one *Participant* record at a time. The **Apply Data List** functionality allows you to fetch the *Specimen* for several *Participant* records in a single operation. To do this, perform the following steps:

1. Fetch the data for the *Participant* category.
2. Fetch the related data from the category *Specimen* for a record of *Participant,* for example, *participant_1*125
3. Add the *Specimen* record to the data list.
4. From the **View Results** screen for the *Participant* records, (see *Figure 2.23. View Search Results*) select *Participant* records for which you want the related *Specimen* records.
5. Click **Apply Data List,** it will fetch *Specimen* for all selected *Participant* records.
6. The status bar displays a success message when the operation is complete. Click **Next** to see the data list. *Figure 2.26* show the resulting data list.

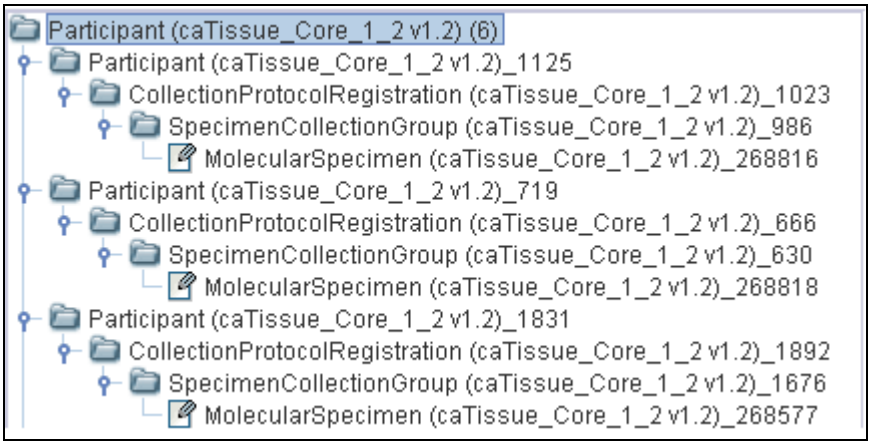


Figure 2.26. Result of Apply Data List to participants with id 1125, 719, 1831

How Apply Data List functionality works

To understand this functionality better, consider a slightly more complicated data list below:

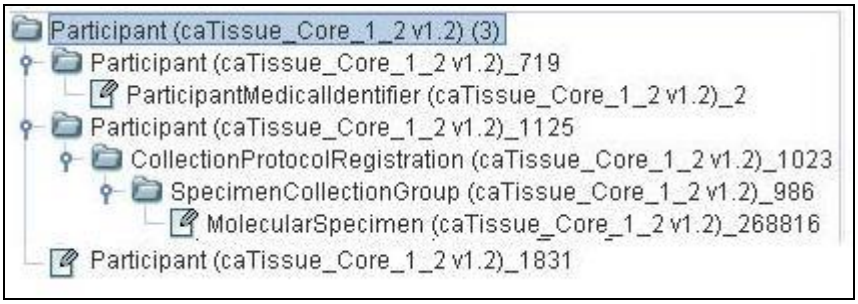


Figure 2.27. Data List

The data list has the following records:

- For *participant_1125*, related *Collection Protocol Registration*, Specimen *Collection Group* and *Molecular Specimen* records are present.
- For *participant_719*, related *Participant Medical Identifier* is present.
- For *participant_1831,* there is no other related data.

Now, **Apply Data List** to the *participants with id 1125, 719, 1831*. *Figure 2.28*. shows the resulting data list

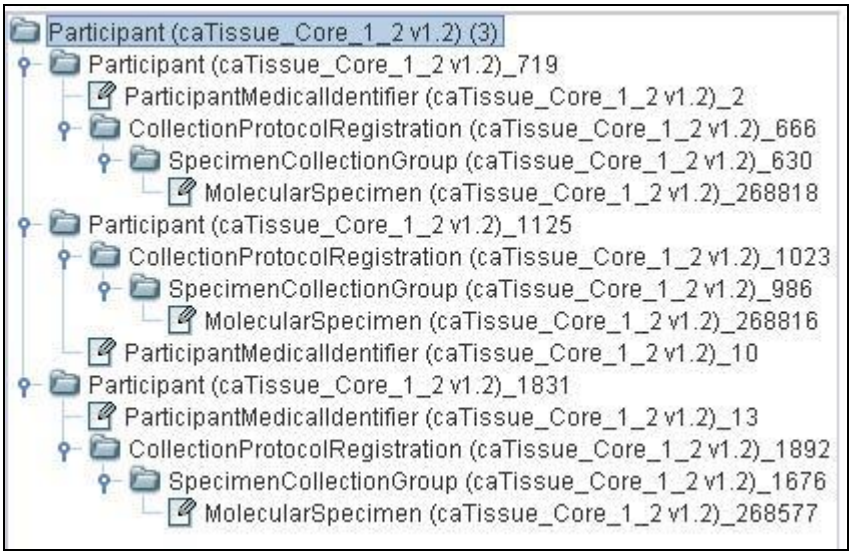


Figure 2.28. Apply Data List

As seen in *Figure 2.28*, application has fetched following related data:

- For *participant _719*, the system fetches *Collection Protocol Registration*, *Specimen Collection Group,* and *Molecular Specimen* (as participant_1125 had *Collection Protocol Registration*, *Specimen Collection Group,* and *Molecular Specimen*).
- For *participant_1125*, the system fetches *Participant Medical Identifier*, as these were the related data for participant_719.

- For *participant_1831*, the system fetches all of *Participant Medical Identifier*, *Collection Protocol Registration*, *Specimen Collection Group,* and *Molecular Specimen*.

As a result, each participant contains the related data for *of Participant Medical Identifier, Collection Protocol Registration, Specimen Collection Group,* and *Molecular Specimen.*

Thus, the **Apply Data List** operation results in a homogenous structure in the data list. The system fetches related data for all the other records such as the related data present for one record.

The Next, Previous and Save Query Buttons

- Click **Next** to view the contents of the data list. It will take you to *Data List* tab.
- Click **Previous** to go back to previous tab, which is *Define Search Results View* tab.
- Click **Save Query** to save the query (refer to *Save Query* above).

Group Results by Service URL

You can configure multiple services to fetch the results. Results are fetched from all the configured services. The View Results page shows all the results from the various services. You can view the results fetched from a particular service by filtering the results. To filter the results follow the steps below:

1. Expand the **Results From** drop down in the *View Search Results* tab.
2. It shows the number of results fetched from each service.

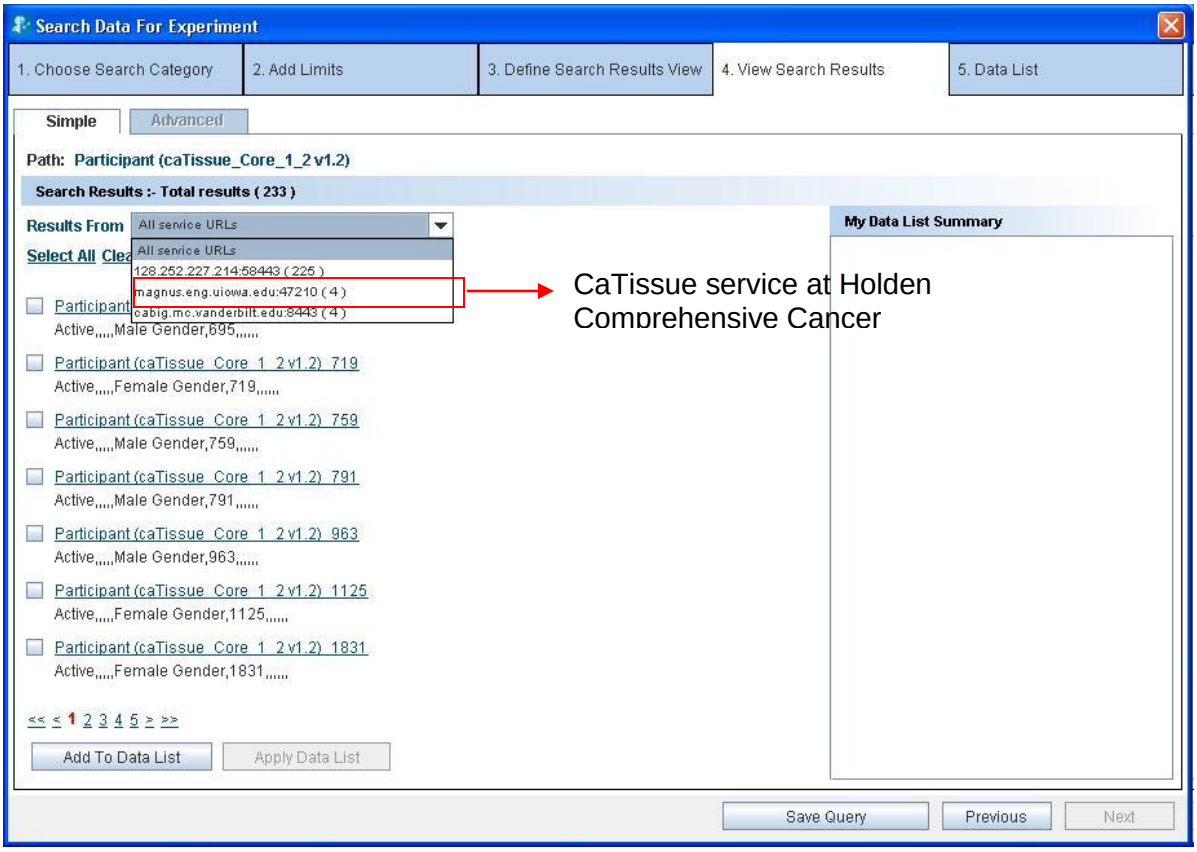


Figure 2.29 Results grouped by service URL.

3. Click on any one of the service in the drop down to see the results from that particular service.

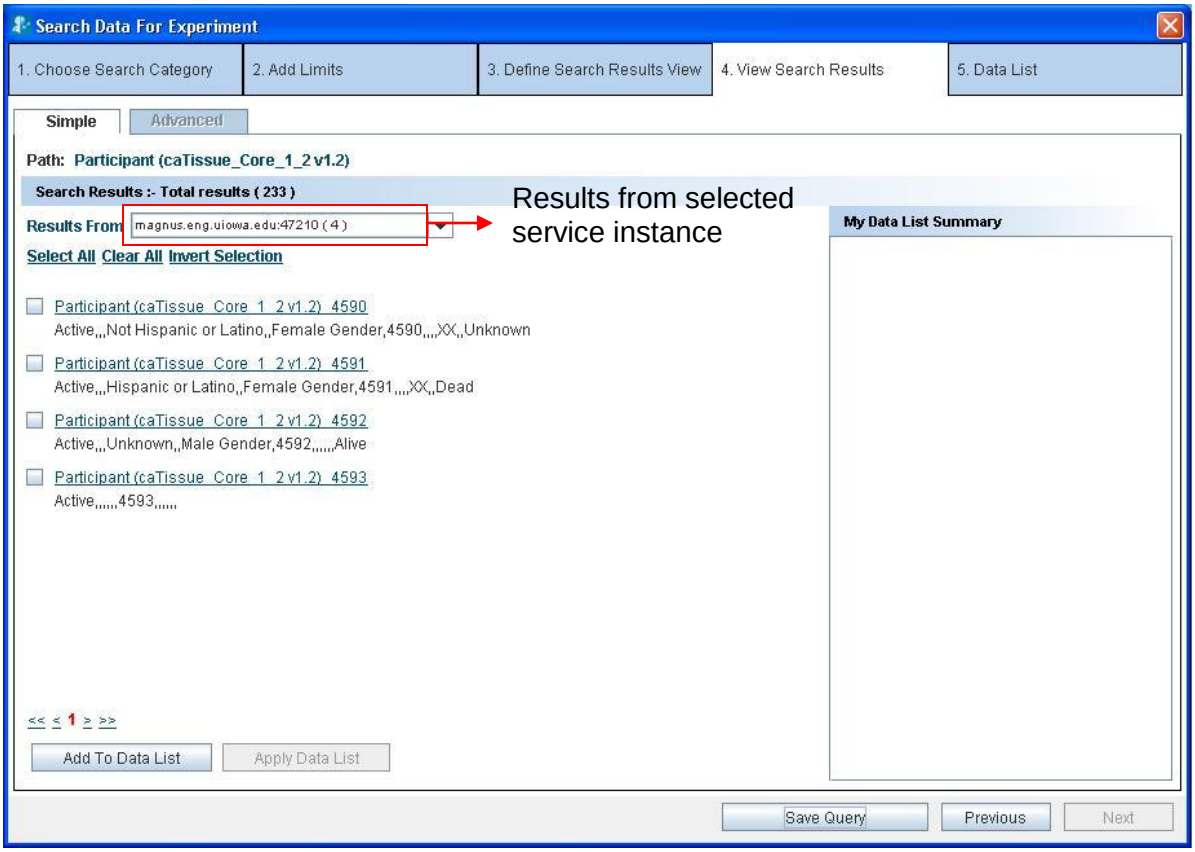


Figure 2.30 Results from the selected service URL instance

You can then proceed with related data and saving the data list for the particular required service rather than all the services.

Viewing and Saving Data List

After adding the records of interest to the data list, you can get the consolidated view of the data list that shows all the records added to the list in a hierarchical tree fashion. These include all the records as well as the details of a particular record. You can also export the data list into a Comma Separated Value (CSV) file or create an Experiment by using the contents of the data list.

The following figure shows the data list tab:

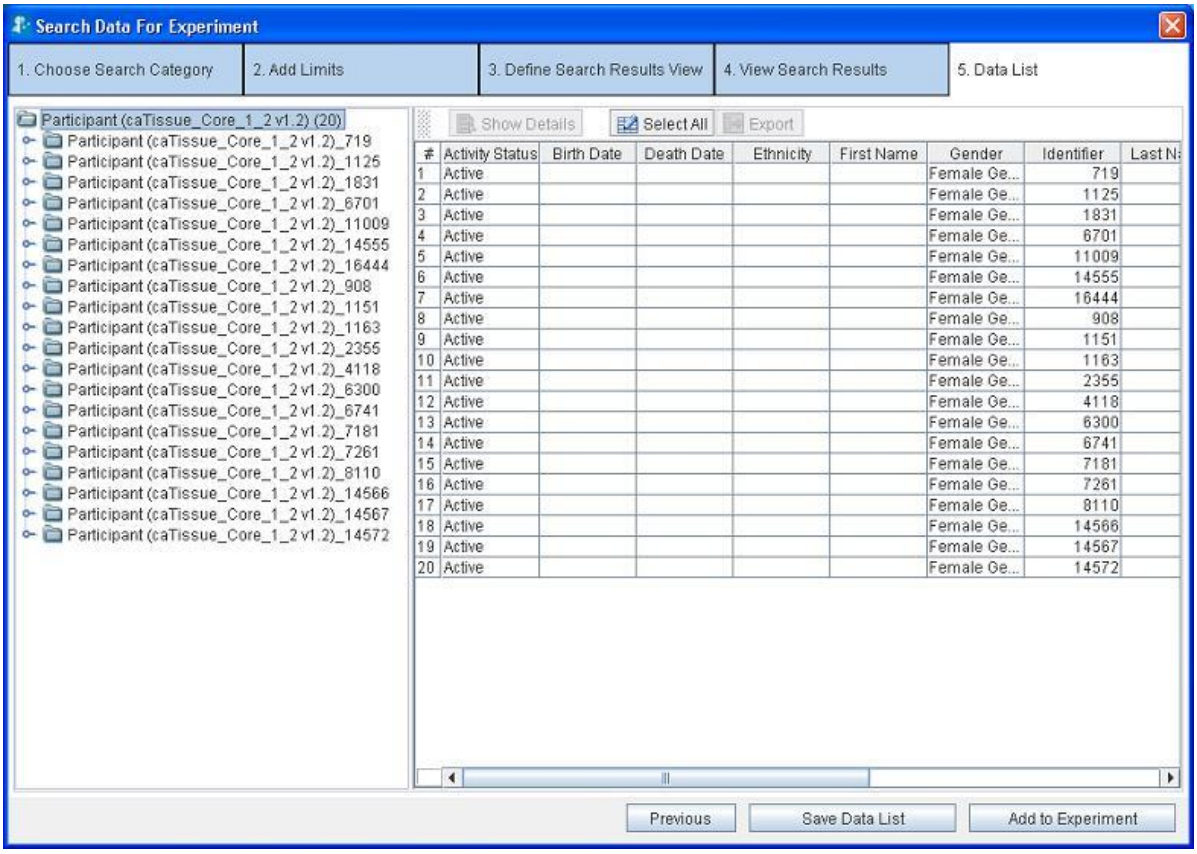


Figure 2.31. Data List

The left-hand side panel shows all the records that you have added to the data list in a tree format. Each node in the tree represents either a category or a record of a category.

- For a node representing a category:
  - The number of records for that category is shown in parentheses. For example, Participant (20) means that you have added 20 records for the Participant category.
  - Clicking on the node updates the right hand side panel to show a spreadsheet containing the records of that category (see figure above) with the columns depicting the attributes and the rows are depicting the records (refer to, *Figure 2.31. Data List*).
- For a node representing a record
  - You see the unique identifier for that record. For example, Participant_719 represents the record with identifier 719.
  - Clicking the node updates the right hand side panel to show the details of that record.

The right hand side panel shows the record details. The details shown in the panel depends on the record chosen in the left hand side panel.

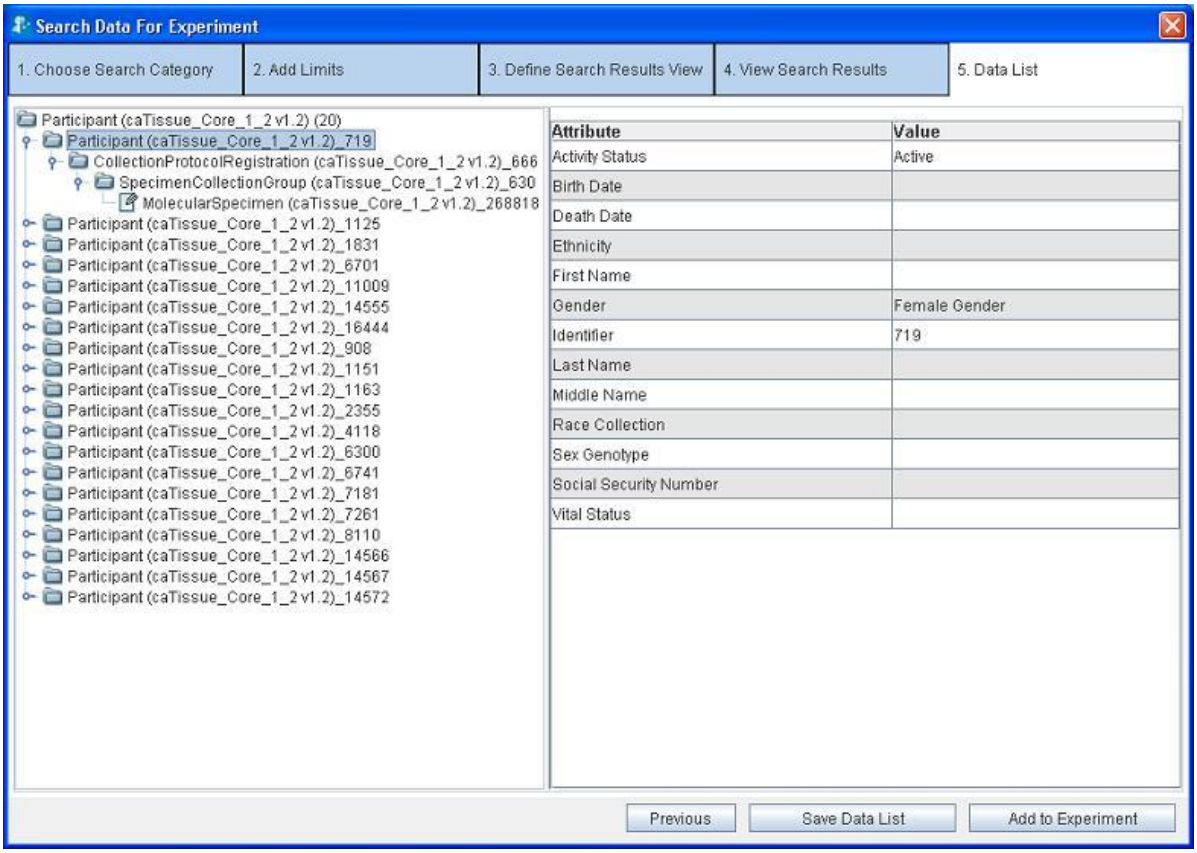


Figure 2.32. Record Details for a single record in Data List

Operations on the Data List

You can perform the following operations on the data list:

- Export the data list into a CSV file – Click **Export**. (see *Figure 2.31. Data List*)
- View details of a particular record – Click **Show Details**.
- Save the data list – Click **Save Data List**.
- Add the data list to an Experiment – Click **Add to Experiment** (see Chapter 3 for additional information).

To save the data list:

1. Click **Save Data List**.
2. Type the title and description for the data list.

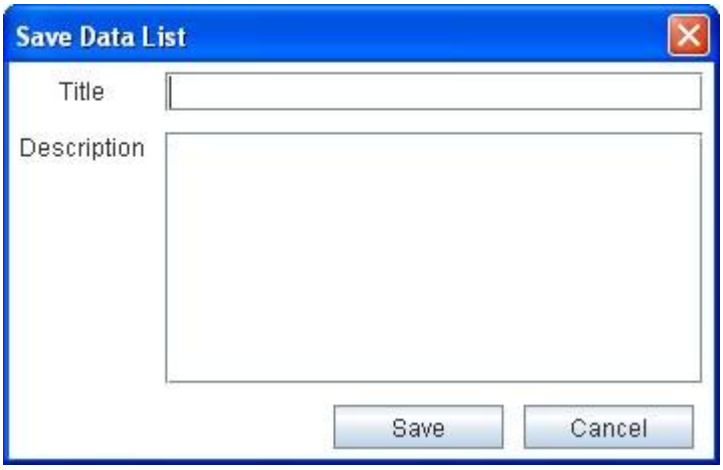


Figure 2.33. Save Data List Panel

3. Click **Save**. When you save the data list, the system displays a success message in the status bar. Application saves the data list in the database.

	Currently, Client Application does not support saving a data list with Date attribute.

Chapter 3 Experiments

After you search and add some data of interest to the data list, you can create an experiment using that data. This chapter explains the various operations that you can perform on the data sets collected.

This chapter covers the following topics:

- *What are Experiments and Projects?*
- *Creating an Experiment*
- *Viewing the Existing Experiments*
- *Working with an Experiment*
- *Filtering Data*
- *Visualizing Data*

What are Experiments and Projects?

An experiment is a collection of data. You can perform the following operations after you create an experiment:

- Visualize data using various viewers such as bar charts and scattered plots.
- Create new data sets by applying filters on the existing datasets

A project is a collection of experiments. A project can contain many experiments or other sub-projects. The grouping of experiments into projects helps you to manage multiple experiments easily.

You can perform the following using caB2B

Step no.	***Description***	***Related Section***
1	Search for data	*Search data – The Query Wizard*
2	Create experiment with this data	*Creating an Experiment*
3	Open the experiment you have created	*Opening an Experiment*
4	Filter the data further if desired	*Filtering Data*
5	Save the filtered data	*How to save filtered data*
6	Plot data as a scattered plot	*Visualizing Data*

Creating an Experiment

To create an experiment, you need to have a saved data list with the required data. See *Viewing and Saving Data List* for more details.

After you have saved the data list, do the following to create an experiment:

1. Click **Add to Experiment** (*Figure 2.31. Data List*). This opens the **Create New Experiment** window.

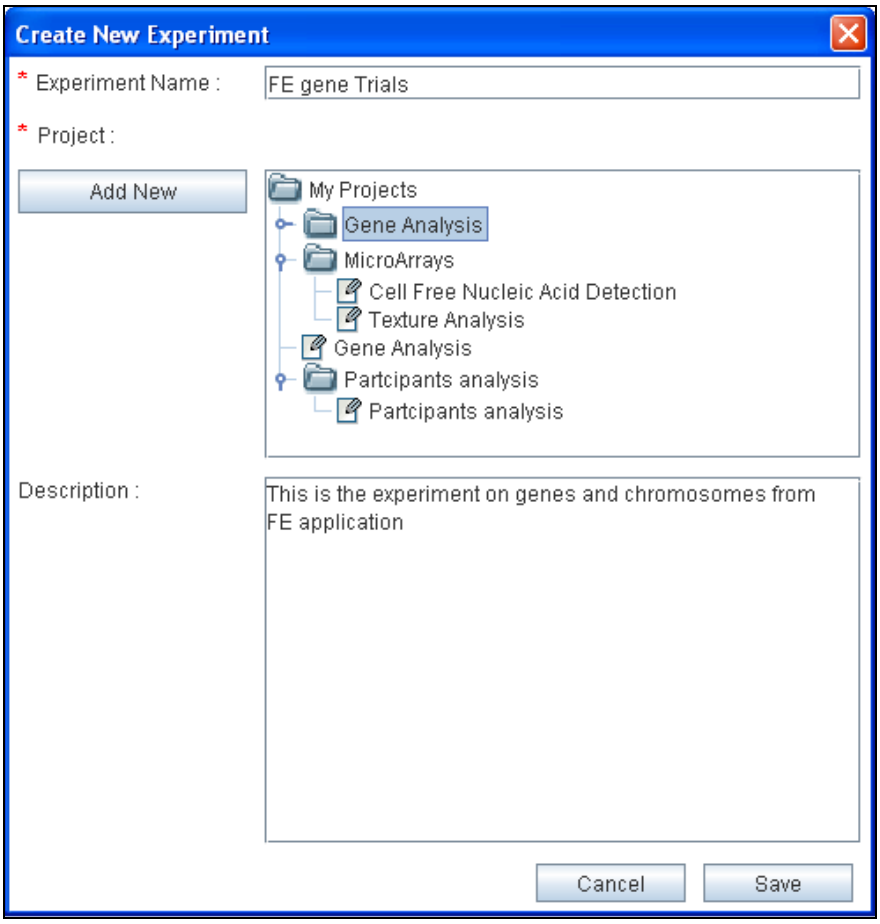


Figure 3.1. Create New Experiment

2. Give an appropriate name to the experiment.
3. Select the project in which you wish to create this experiment (see note below).
4. Provide a description for the experiment.
5. Click **Save**.

You can now close the **Search Data** wizard to return to the **Home** page and to open the experiment.

Similarly, you can search for gene, RNA, protein and save the experiment. Further, the features of experiment are explained by taking an example of gene and related intramodel associations.

- To create a new project, perform the following:
  1. Select the parent project under which you wish to create a new project
  2. Click **Add New**. This adds a new sub project under the selected project with the default name– New Project.
- To rename the new project
  1. Select the project.
  2. Press F2.
  3. Type the desired name.
  4. Press ENTER.

Viewing the Existing Experiments

On the caB2B home page, click the **Experiment** tab. This opens the My Experiments page.

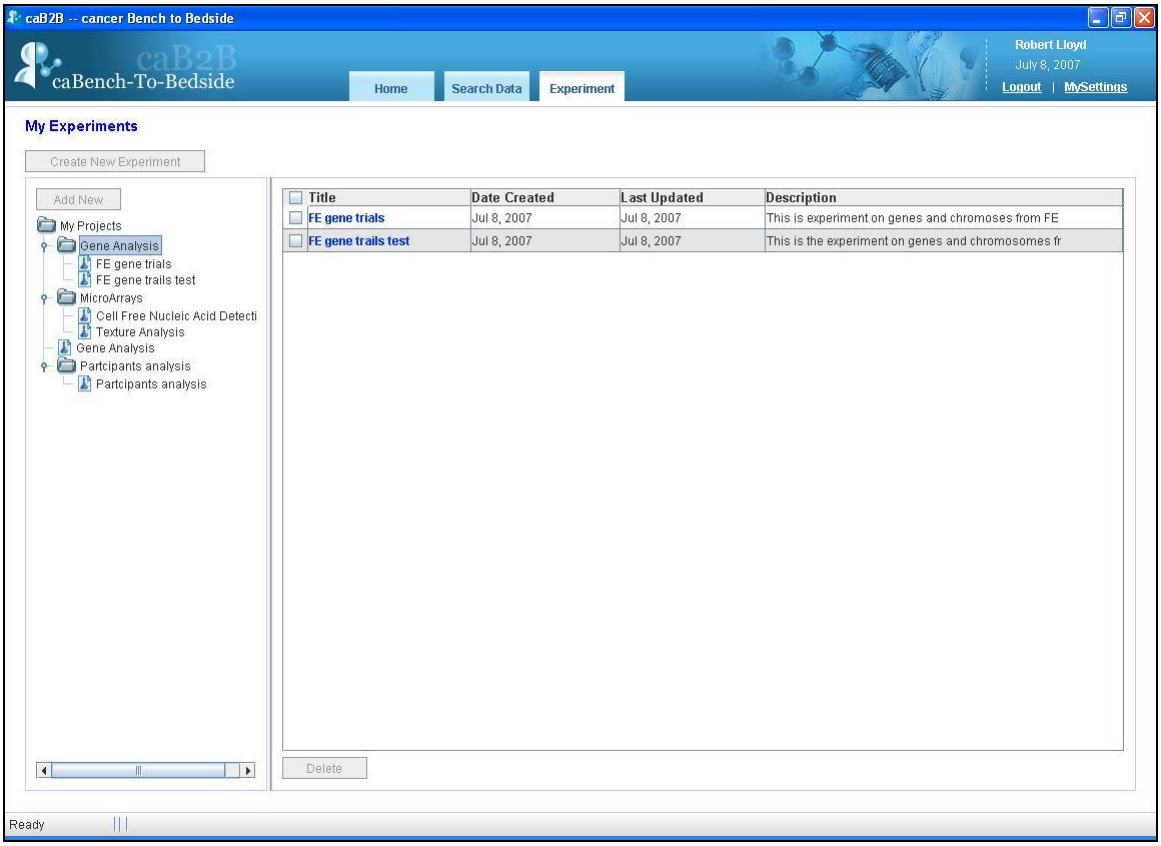


Figure 3.2. My Experiments

You can see all the projects and experiments that you created in the left panel.

When you select a project in the left hand side panel, the system displays a list of experiments in the project in right hand side panel.

Opening an Experiment

To open an experiment, go to the My Experiments page and click on the experiment title seen in the right panel (see *Figure 3.2. My Experiments*).

On opening an experiment, you see the **Experiment Details** page. The experiment details page shows the data categories added to experiment. It gives you an option to modify the experiment by adding a new column, adding new data, exporting data etc. All these options are also available on the right click menu.

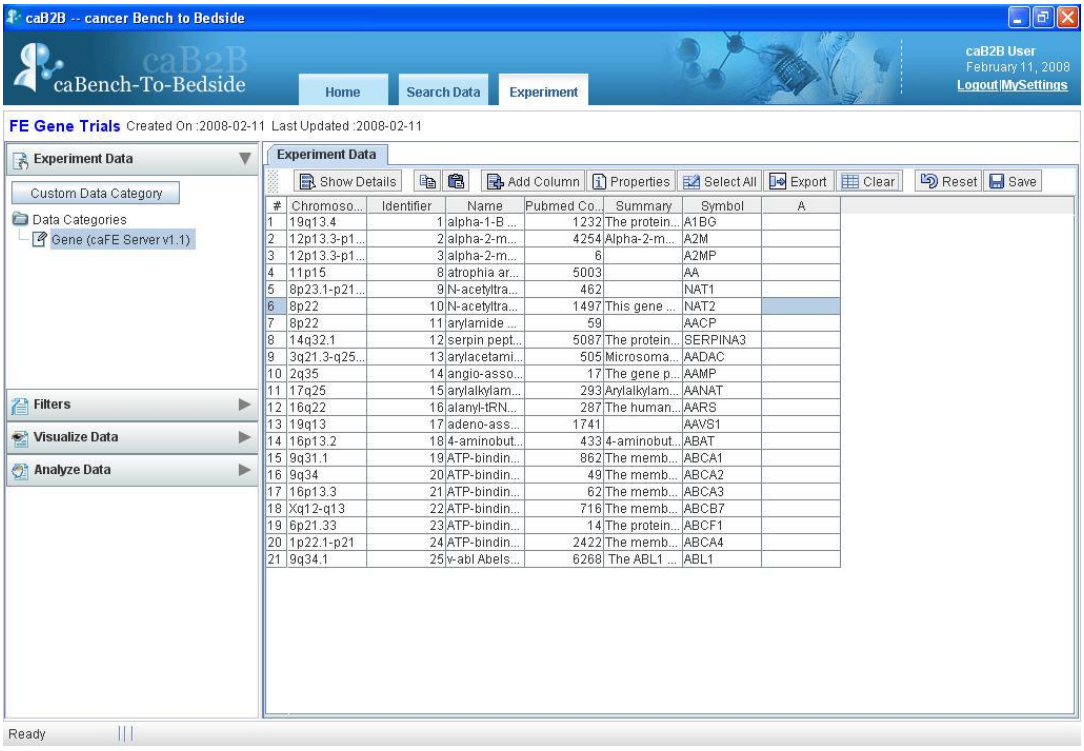
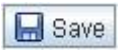


Figure 3.3. Experiment Details

The table below lists the various options provided by the spreadsheet on the experiment details page:

Button	Description
Show Details	Shows details of the particular record
	Copies data of the selected cells
	Pastes the data to the specified cells
Add Column	Adds a new column to the spreadsheet
Properties	Shows the properties of the spreadsheet, like which attributes are visible, what filter is applied
Select All	Selects all the records

Button	Description
	Exports data to a CSV file
	Clears the user added data from the selected cell
	Undo's all the changes and shows the original data
	Creates a data category

Working with an Experiment

The **Experiment Details** page consists of the following sections:

- The **Operations** panel on the left hand side
- The **Visualization** panel on the right hand side

The Operations panel

The **Operations** panel encompasses the following collapsible boxes.

Table 3.1. Operations on the Experiment Page

UI Elements	Description
	It displays a list of all the data categories present in the experiment. See *How to save filtered data* in this chapter for more information.
	It shows all the filters that you applied to the data categories. If none, then it shows filters that you can apply to the data of the selected data-category. See *Filtering Data* in this chapter for more details.
	It shows the visualization options available for the selected data-category. See *Visualizing Data* in this chapter for more details
	Shows the analytical services that are applicable to the selected data-category

Visualization panel

This is the right hand side panel on the **Experiment Details** page (*see Figure 3.3. Experiment* Details). This panel can contain several pages at a time (see figure below). You can switch between the pages by clicking on the **tabs** at the top of this panel.

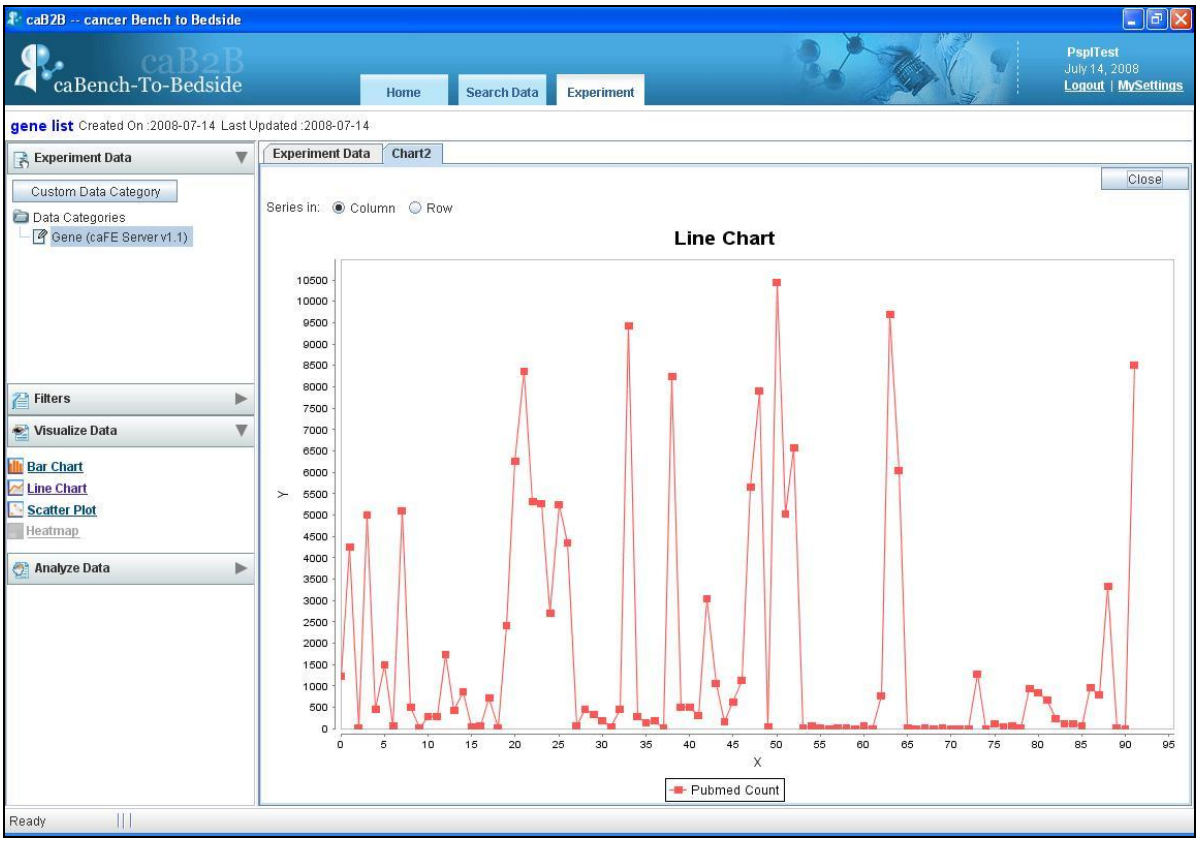


Figure 3.4. Visualization Panel

The visualization panel updates itself based on the operations you perform in the operations panel. For example,

Operations panel	***What happens in Visualization pane***
Click a data-category in the Select **Data Categories** box	The **Experiment Data** tab is updated to show the records of that data category
Choose a visualization from the **Visualize Data** panel	A new tab is created to display the desired visualization

Filtering Data

The purpose of creating experiments is to perform analysis on the data. To aid the analysis, caB2B provides features to filter the data so that you can zero in on the data of interest.

After filtering the data, you can save the filtered data as a **Data Category**. For example, if you have created an experiment with all female participants, you can now create different data categories based on age less than 50, age between 50 and 65 and age greater than 65.

How to filter data

1. Select the column to be filtered, select the filter from the **Filters** stack box, (located in the left hand side panel)

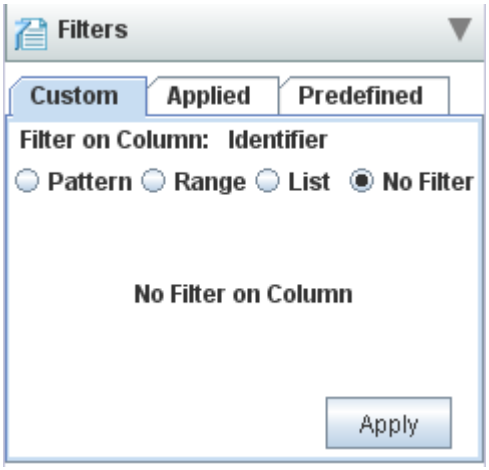


Figure 3.5. Filter Stack box

2. System displays appropriate filter controls.
3. Choose one of the available filters and specify the conditions that the data must satisfy (refer to, *Appendix D: Types of Filters*).
4. Click **Apply**.
5. The **Experiment Data** tab shows the filtered data.
6. The filter that you have applied is also shown in the **Applied** tab of the filters stack box on the left hand side.

You can now save the filtered data as a new data-category (refer to *How to save filtered data* for more information) and use it to perform analyses or visualization.

	You can filter on multiple columns by successively applying filters on those columns. For example, to obtain *female participants between age 30 and 40* 1. Apply the filter “Gender = female”. 2. Then apply the filter “age between 30 and 40”. For the details of the available filter controls and their mechanism of action, see *Appendix D: Types of Filters.*

How to edit a filter

1. Choose the column to edit the filter.

The Filter stack box shows an appropriate filter with the values that you had specified earlier.

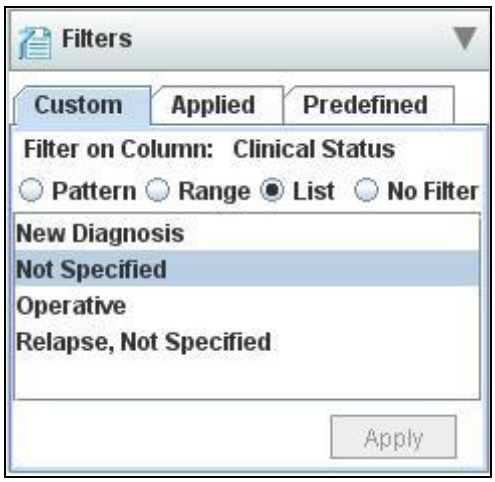


Figure 3.6. Edit Filter

2. Specify new values for the filter (refer to, *Appendix D: Types of Filters*).
3. The **Experiment Data** tab shows data based on the new filter.
4. The filter criteria shown in the **Applied** tab of filters stack box in the operations panel is also updated.

How to remove a filter

1. Choose the column whose filter you wish to remove.
2. Select the **No Filter** option.
3. Click **Apply.**
4. The **Experiment Data** tab shows data after removing the filter.
5. The filter is also removed from **Applied** tab in the filters stack box in the operations panel.

How to save filtered data

Once you apply filter to a data set, you can save the filtered data as a new **data category**. You can create as many data categories as you want. You can use these data categories for data analysis.

To save a filtered data category,

1. Click **Save** (It is located in the toolbar of the right hand side panel of the experiment details page. Refer to *Figure 3.3. Experiment* Details)
2. The **Save as Data category** panel appears.
3. In the **Title** field, type the title of the new data-category.

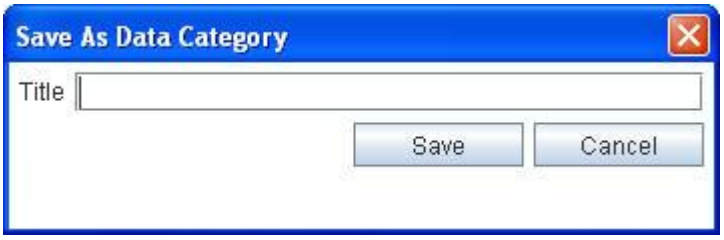


Figure 3.7. Save Data Category Panel

4. Click **Save**.

The new data category will appear under **Custom Data Categories** section in **Experiment Data** box of the operations panel.

Custom Data Category

In an experiment, there are various data categories, like participant, specimen etc. An individual data category does not make much sense. Data in different data categories is related to each other. Thus, a user would like to combine data from different categories to generate a unified view for all the categories.

Custom Data Category (CDC) is a way to bring data from various categories under a single category. For example, you want to see the specimen details and participant information in a single view. For this, you can create a CDC with the desired attributes from the participant and the specimen class and create a new category.

How to create CDC

1. Click **Custom Data Category** (located at the top of the Experiment Data box on the left hand side panel of the experiment page).

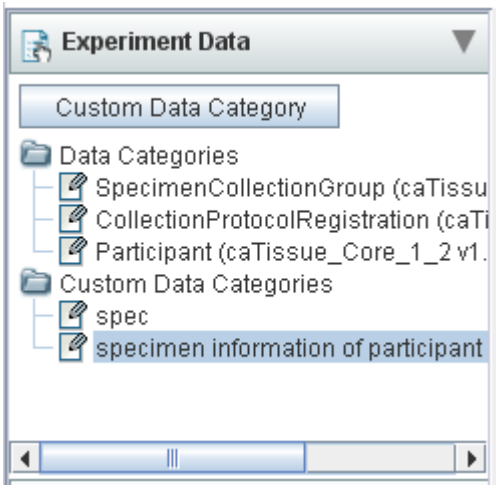


Figure 3.8. Custom Data Category

The **Create Custom Data Category Panel** appears.

2. In the **Custom Data Category Title** field, provide the title.
3. In the **Data List** field, select the data list.

4. In the **Category** field, select the category.
5. Select attributes from **Available Attributes** to be added in CDC.
6. Click **Add/Add All** to add to **Selected Attributes**.
7. Select attributes from the **Selected Attributes** and click **Remove/Remove All** button to remove the attributes from CDC.
8. Click **Save**.

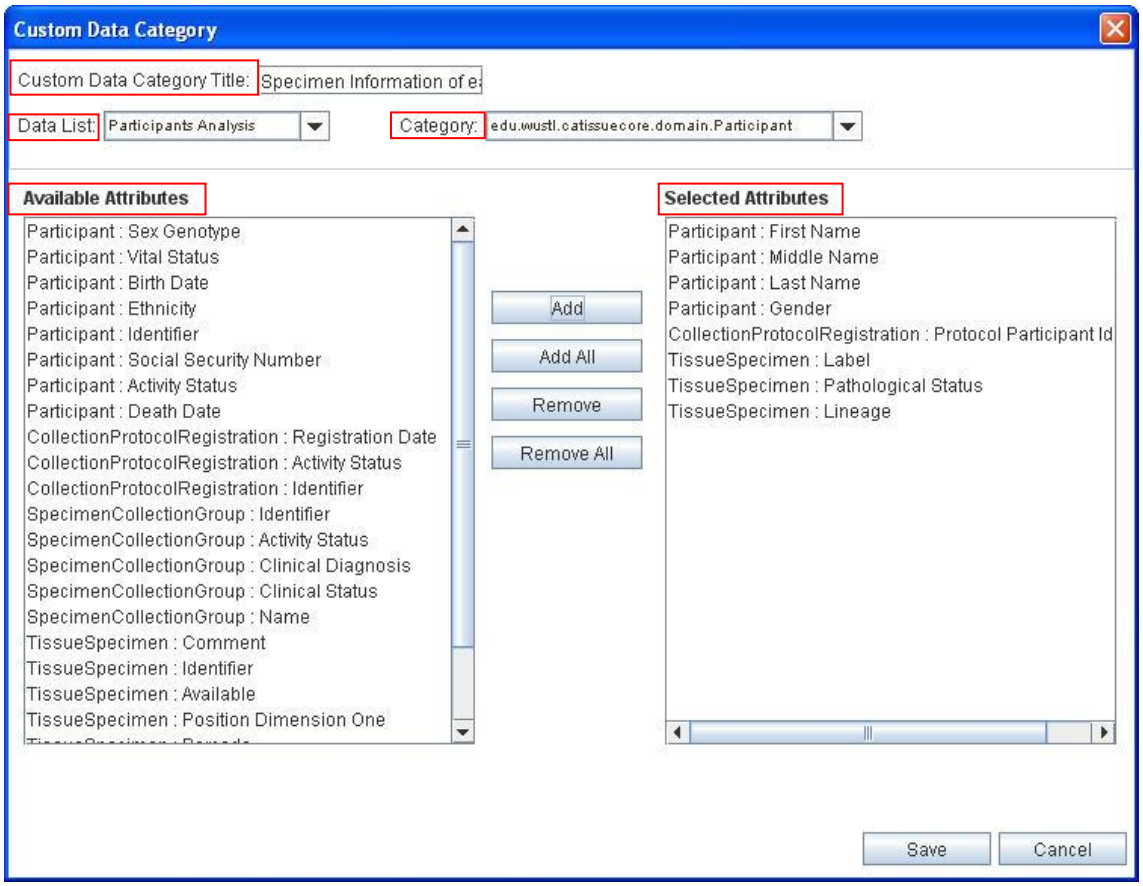


Figure 3.9. Custom Data Category Panel

The new CDC appears under **Custom Data Categories** section in **Experiment Data** box of the operations panel as shown below.

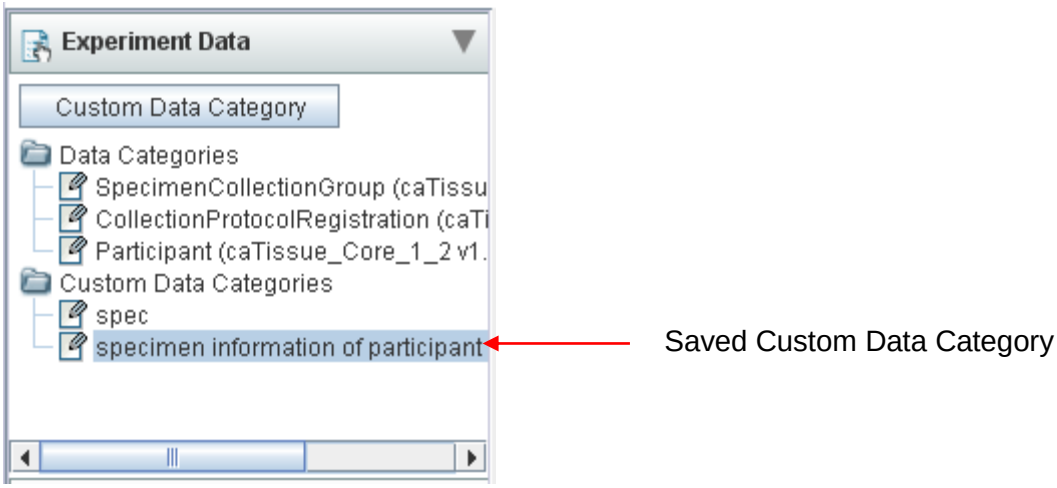


Figure 3.10. Saved Custom Data Category

Visualizing Data

You can visualize the data in the experiment by using various graphical components such as charts and scatter plots. These options provide a graphical summary of the data allowing comparison, variation, and co-relational analysis of the data.

Some terms used:

Data series: Related data points plotted in a chart that originates from data rows or data columns. Each data series in a chart has a unique color or pattern represented by a *Legend*. You can plot one or more data series in a chart.

Legend: A small color box that identifies each series in the chart

The current version of caB2B supports following visualization options in **Visualize Data** box.

Visualize option	*Description*
Bar Chart	This is a chart with rectangular bars of lengths proportional to the data value it represents
Line Chart	This is a chart with a line joining the points that represents data value
Scatter Plot	A scatter plot will show a linear or non-linear relationship between the variables (column selected). It also helps in finding outliers in the data.

Table 3.2. Types of charts supported

How to plot data series

1. Select the data that you want to plot from **Experiment tab**.
2. Click on the link for type of the chart you wanted from the **Visualize Data** Stack Box, for example Bar Chart
3. It will open the chart in a new tab in right hand side panel.

4. By default, data series are from columns on the **Experiment Data** tab. To change the data series toggle the radio buttons in the upper left corner of the chart area.

Following images represents the various charts for **PubMed** count for few Genes obtained from FE server.

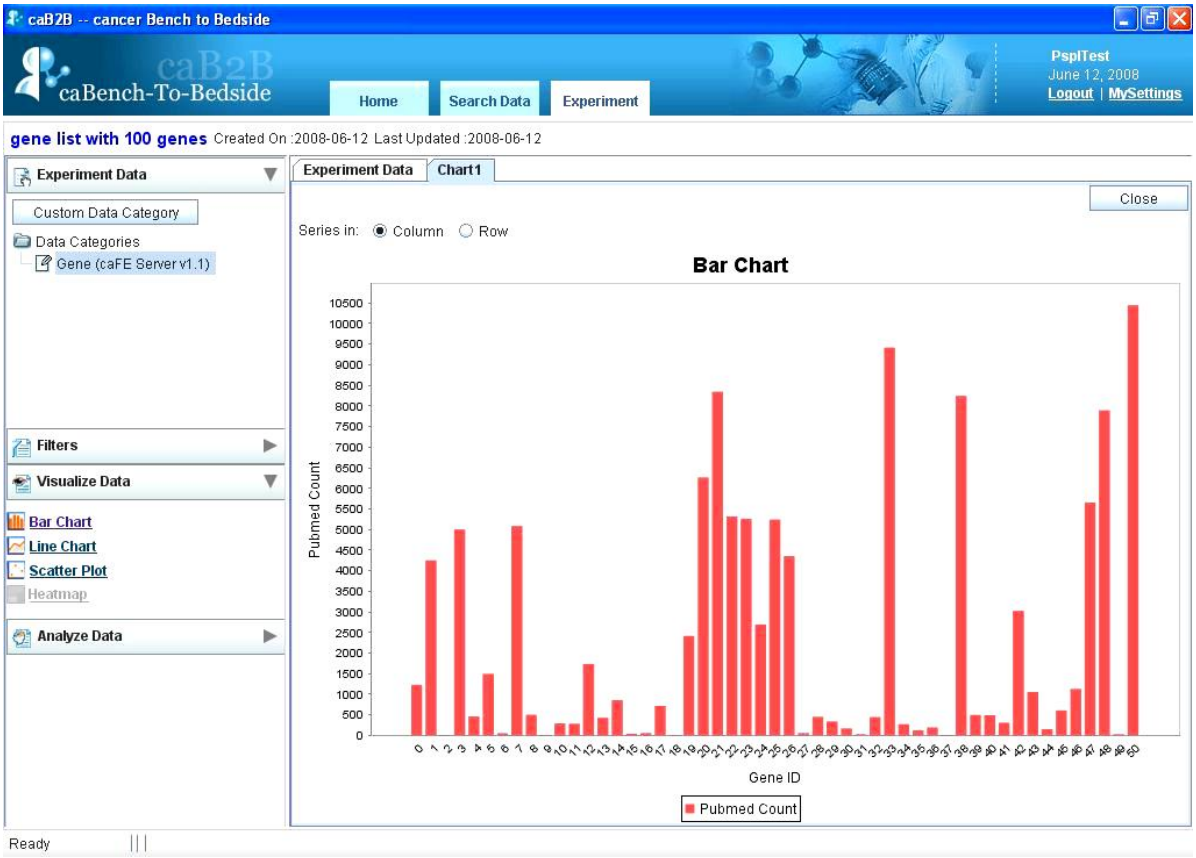


Figure 3.11. Bar chart example

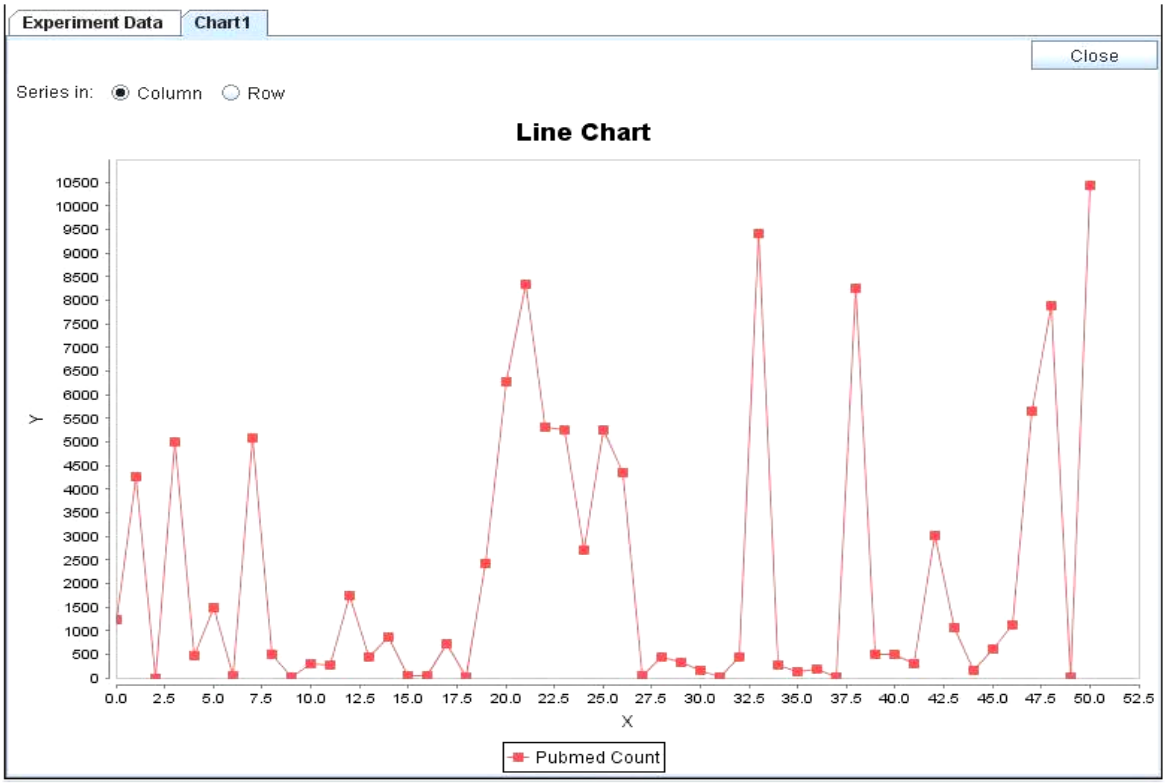


Figure 3.12. Line chart example

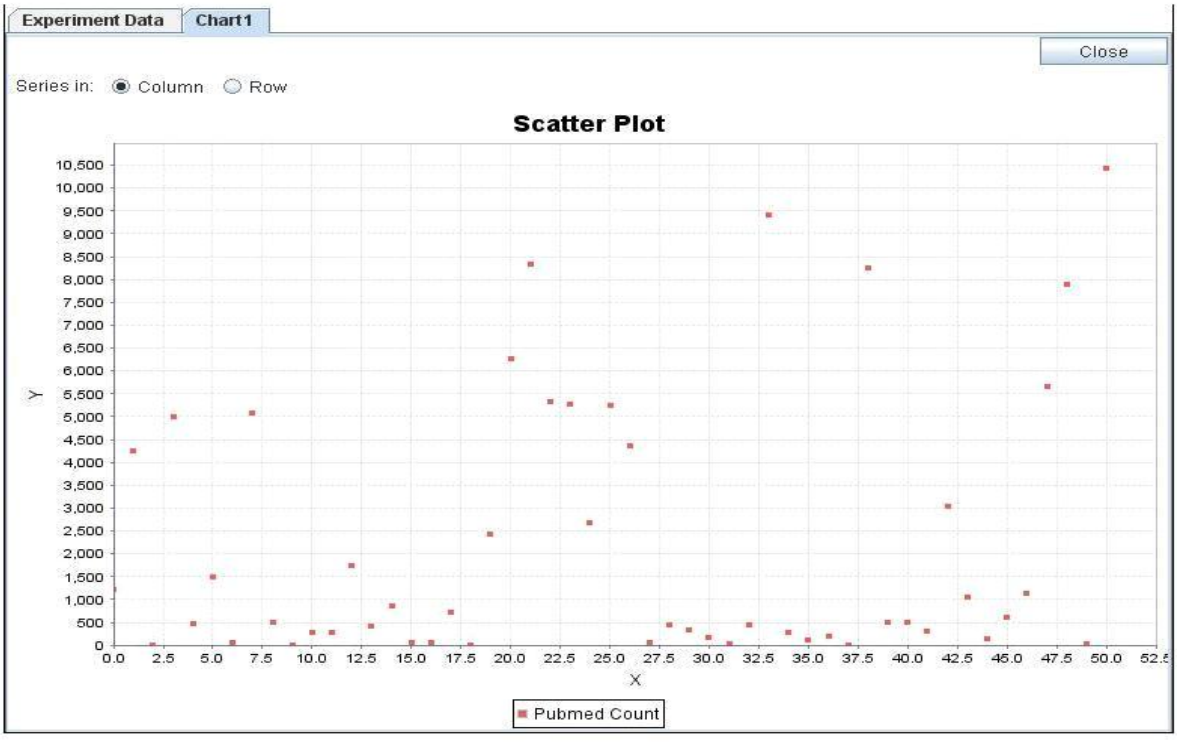


Figure 3.13. Scatter Plot example

How to change data series from rows to columns and vice versa?

You can configure whether data series are from rows or columns by selecting appropriate **Series In** options.

- To have the data columns to represent the data series, select Column option.
- To have the data rows to represent the data series, select Row option.

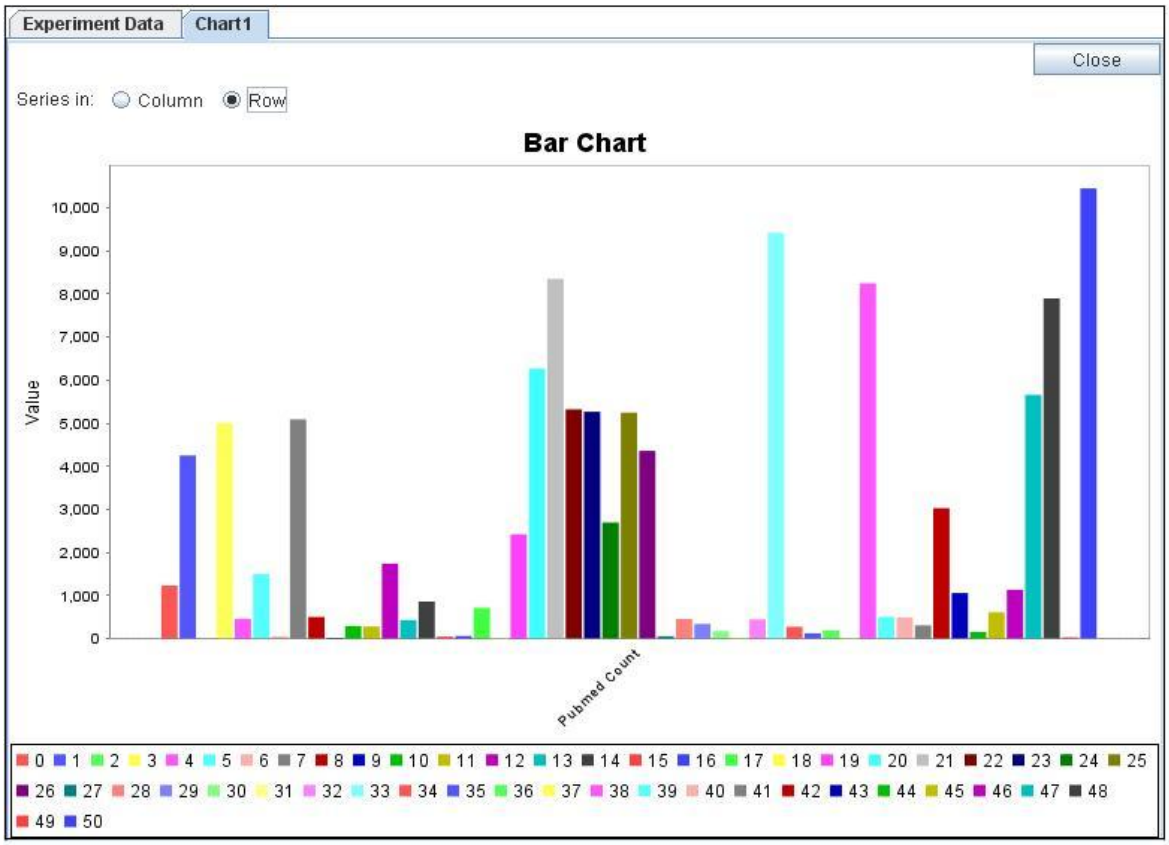


Figure 3.14. Bar chart with data series from rows

FYI.	You can change the type of the chart that you are currently viewing by selecting required chart link from **Visualize Data** box. You can create multiple charts for the data. The name of the tab added for each chart is “chart” followed by a unique number for that chart. For example chart1, chart2, etc.

Appendix A Examples of category search

This appendix provides examples of the advanced search strategies that are possible in caB2B, while searching for categories.

Text-based Search - For category name information

To search categories having “gene” in their name do the following:

1. In the **Search** box, type ***gene***.
2. Expand the **Advanced Search** box
3. Select the **Category** check box.
4. Click **Text**.
5. Click **Search**.

The results of this search are as shown below. All the categories shown below have *gene* in their names.

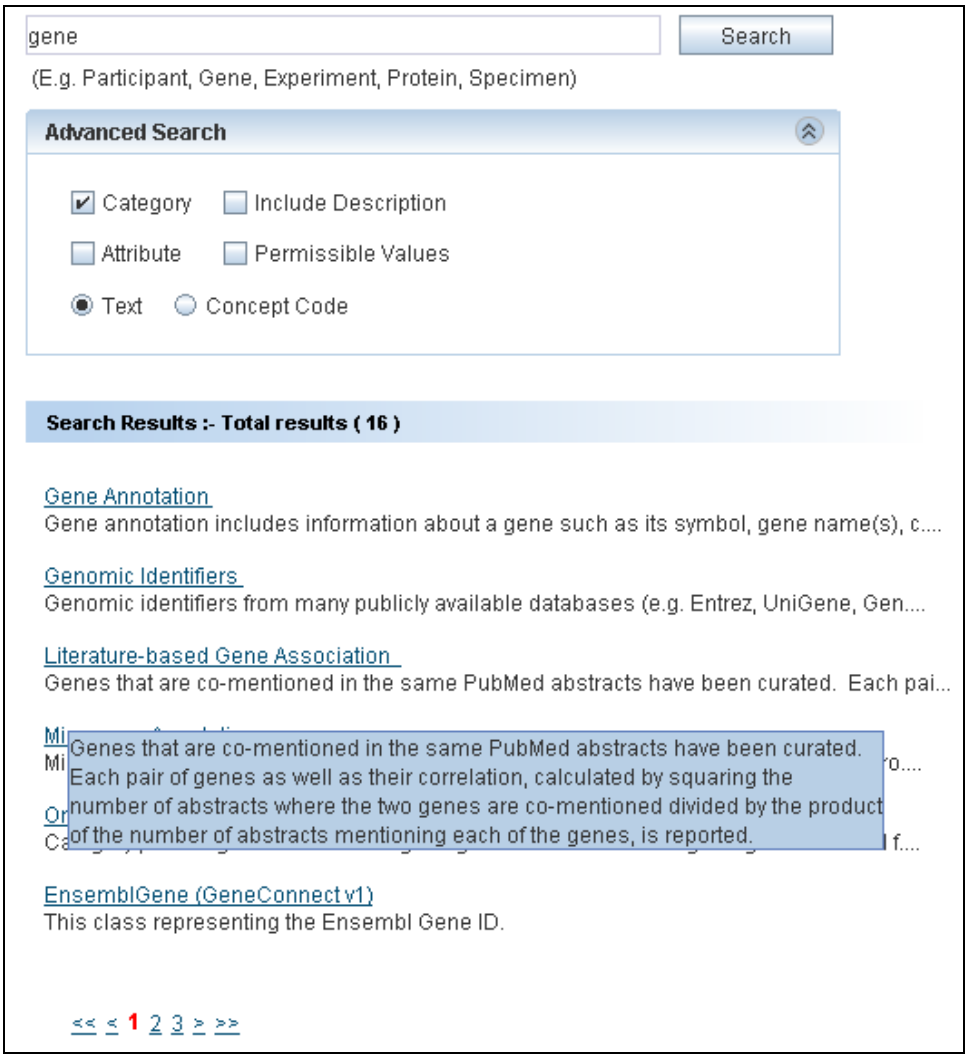


Figure A.1. Search Result of Example 1

Text-based Search - For category and attribute name and description

To search the categories having “Protein” in their name or description or to search the attributes having “Protein” in their name or description, you need to do following:

1. In the **Search** box, type ***Protein***
2. Open **Advanced Search** box**.**
3. Select the **Category** check box.
4. Select the **Attribute** check box.
5. Select the **Include Description** check box.
6. The **Text** option button is selected by default.
7. Click **Search**.

The result of this search is as shown in the screenshot below.

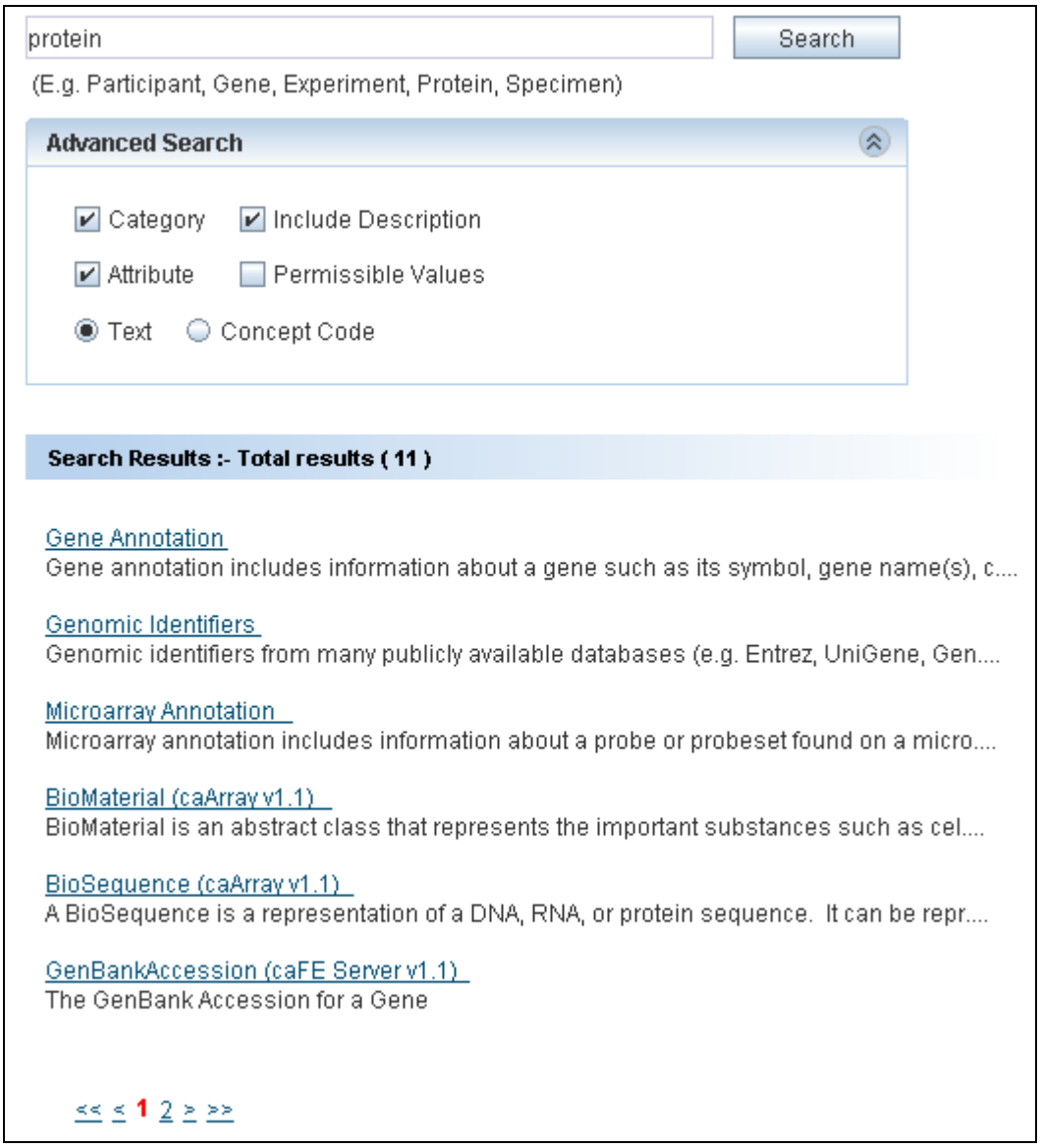


Figure A.2. Search Result of Example 2

Concept Code Based Search - For category name information

To search categories having concept code “c29867”**:**

1. In the **Search** box, type ***c29867***.
2. Open the **Advanced Search** box.
3. Select the **Category** check box.
4. Select the **Concept Code** option button.
5. Click **Search**.

The result of this search is as shown in the screenshot below.

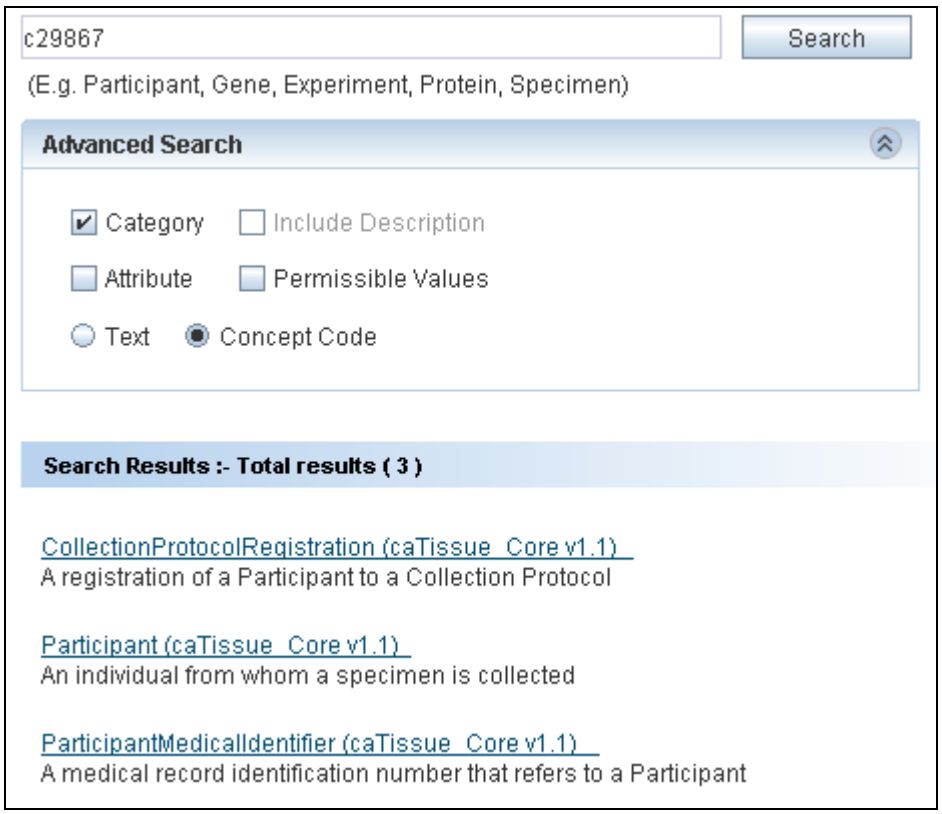


Figure A.3. Search Result of Example 3

Concept code based search - For attribute name information

To search categories whose attribute has concept code “C42614”:

1. In the **Search** box, type ***C42614***.
2. Select the **Attribute** check box.
3. Select the **Concept Code** option button.
4. Click **Search**.

The result of this search is as shown in the screenshot below.

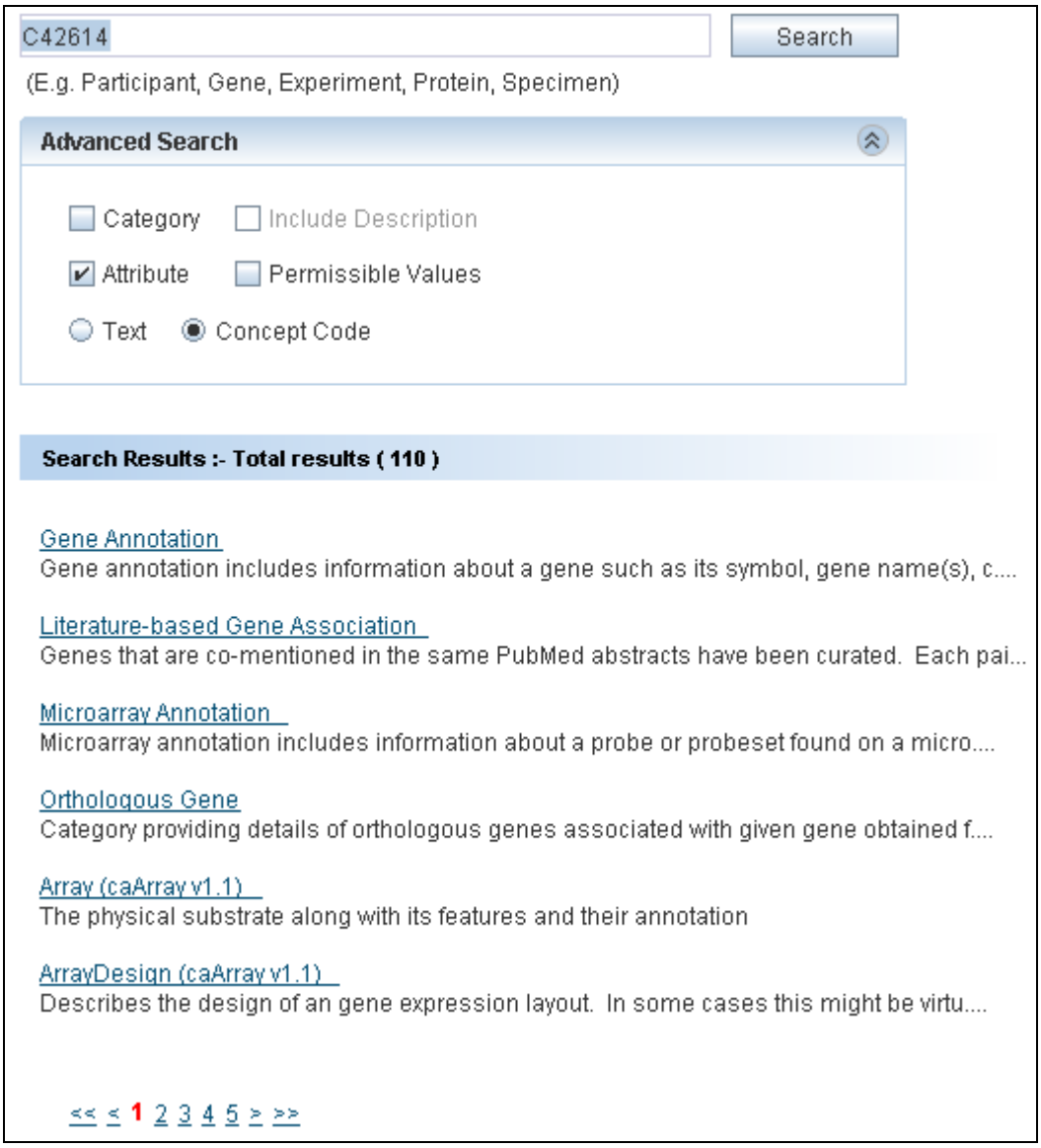


Figure A.4. Search Result of Example 4

Appendix B Define Limit: Operators and Values

This appendix explains the various operators. It also explains how to specify values while defining limits on a category.

The following table explains the different operators and provides examples of the values to be specified:

Operator	Data types for which applicable	Sample condition	Expected Results
Equals	Numeric Date String	Name **equals** "John"	Returns only John and not john
Not equals	Numeric Date String	Identifier **not equals** 32	Returns all records except those having identifier = 32
Less than	Numeric Date	Participant's birth date **less than** 2000/1/1	Returns all Participants born before January 1st, 2000
Less than or equal	Numeric Date	Signal value **less than or equals** to 0.5	Returns all records for which signal value is not more than 0.5
Greater than	Numeric Date	Cell count is **greater than** 10000	Returns all records with Cell count > 10000
Greater than or equals	Numeric Date	Specimen's collection-date **greater than or equal to** 2005/5/1	Returns all specimens collected on or after May 1st, 2005
Between	Numeric Date	Patient's admission date **between** 2000/1/1 and 2000/12/31	Returns all Patients admitted in the year 2000
Is null	Numeric String Date	Patient's contact number is **null**	Returns all patients who do not have any contact number
Is not null	Numeric Date	Participant' middle name is **not null**	Returns all Participants for whom middle name is specified

In	Numeric String Enumerated	• Signal value **in** 0.1,0.5,0.7, • Name **in** “John”, ”john”	• Returns all records whose signal value is 0.1 or 0.5 or 0.7 • Returns both John and john
Not in	Numeric String Enumerated	Participant’s gender **not in** male, unspecified	Returns all participants whose gender is either not specified or gender is not male.
Starts with	String	Participant’s first name **starts with** Jo	Returns all participants whose first name starts with Jo like John, Jona Does not return the participant with name Scott, Henry, Jen
Ends with	String	Participant’s last name **ends with** n	Returns all participants whose last name ends with n like John, Jon, and Jen Does not return participant with name Scott, Henry
Contains	String	Participant’s first name **contains** ann	Returns all participants whose first name is Joanna, Anna, and Johanna. Does not return participant with name Jen, Denny

- For attributes with enumerated values and Boolean attributes, clear a value by clicking on the value with the CTRL key pressed.
- If you use the **In** operator for a non-enumerated attribute, you can specify multiple values as follows:
  - Separate multiple values using a comma. For example, to specify the constraint **Identifier IN *2, 4, 5, 7***, you should type the value as ***2,4,5,7*.**
- If a value itself contains a comma, you can enclose the value in double quotes. For example, to specify the constraint **Name IN “*Einstein, Albert*”, “*Newton, Isaac*”** the value that you should enter is **“*Einstein, Albert*”, “*Newton, Isaac*”.**
- For attributes that requires a date value, you can click the icon to launch a date selector. You can also type a date string directly in the text box. The format of this string must be **YYYY/MM/DD**. A valid date string becomes green whereas an invalid date string turns red. System will ignore an invalid date string.

Appendix C Record Details View

The details of a record usually consist of the values of the attributes of the category (See *Figure 2.24. Results' Details).* This section provides examples of special categories that result in a slightly different view of the record's details.

Administrator defined categories

The Administrator builds categories by manually grouping relevant attributes from one or more existing categories.

For each record of such a category, you can see the following information:

- Values of the attributes in a table
- Related information as a tree, the leaf nodes of this tree are attributes; you can see the values of those attributes in the tree.

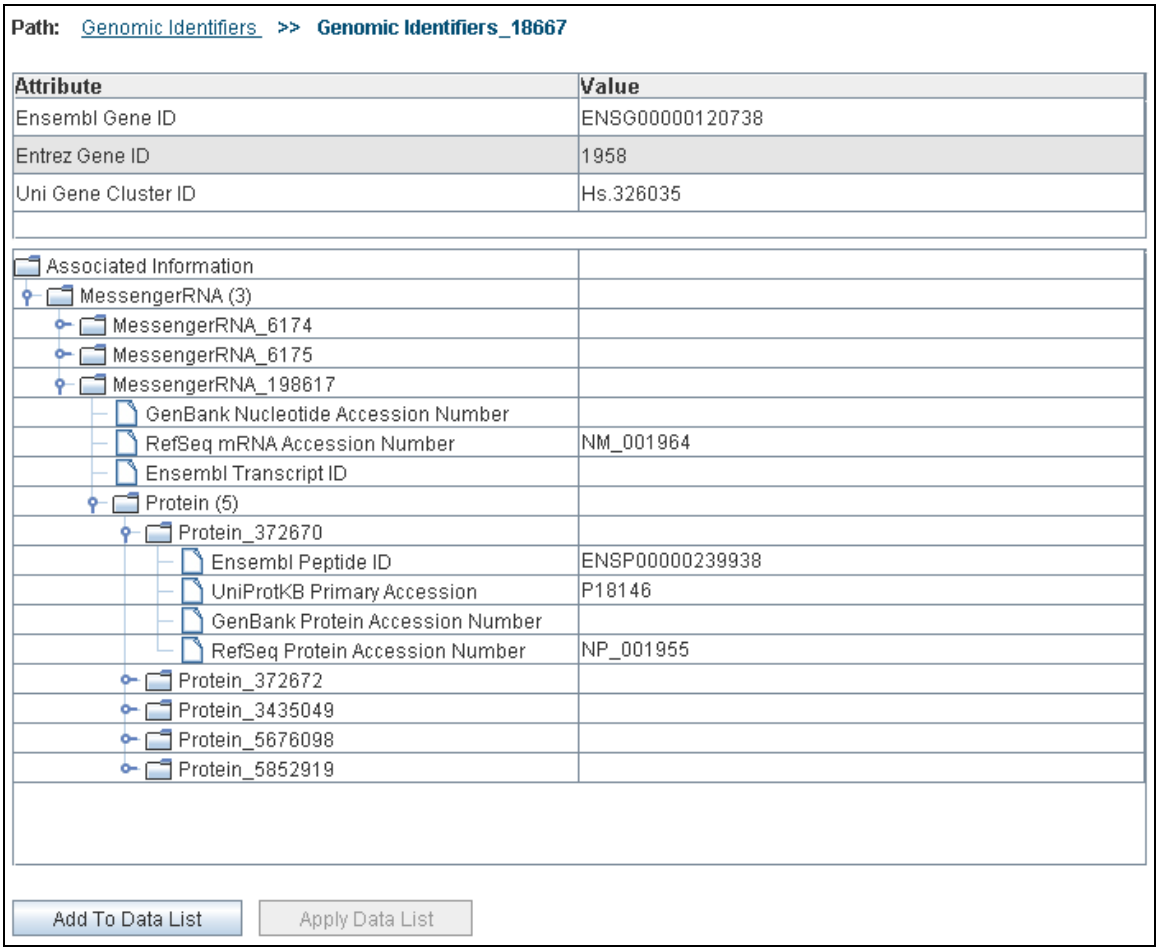


Figure C.1. One record of an administrator defined category (Genomic Identifiers)

Appendix D Types of Filters

A filter is shown when you choose to filter the data of a data-category (see *How to filter data*).

The type of filter depends on the column based on which you wish to filter data.

Table D.1: Column Type and Applicable Filter

Data type of the column	***Applicable filter***
Numeric	Range filter
Text	Pattern filter
Column with permissible values	Enumerated filter
Boolean	Enumerated filter

Using the Range Filter

The system shows a **Range Filter** if the column selected as a basis for the filter has numeric values. It allows you to select the desired minimum and maximum values of the column.

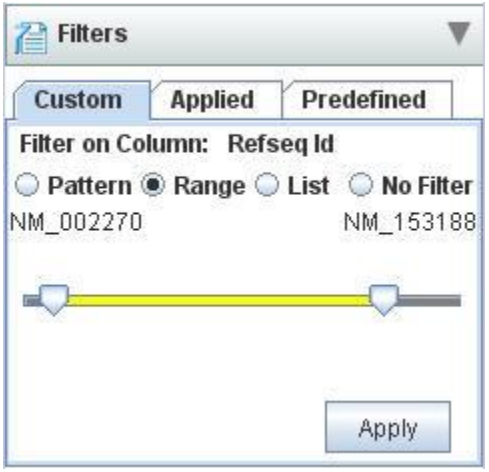


Figure D.1. Set Filter for Range

There are two sliders, the left slider and the right slider. The two ends show the value denoted by each slider.

1. Position the left slider to the minimum value desired.
2. Position the right slider to the maximum value desired.
3. Click **Apply** to apply the filter.

Using the Pattern Filter

The system shows a **Pattern Filter** if the column selected as a basis for the filter has text values. It allows you to specify a pattern that the values of that column should satisfy.

Figure D.2. Pattern Filter

In the text field, type the desired pattern and click **Apply**.

Using the Enumerated Values Filter

The system shows a **List Filter** if the column selected as a basis for the filter has a set of permissible values. It allows you to specify the values that the column should satisfy.

Figure D.3. Enumerated Values Filter for the column gender

Select the desired permissible values and click **Apply**.

Index

Administrator defined category, 67
Advanced search, 15
Ambiguity resolver, 22
Apply Data List, 37
Breadcrumbs, 37
caB2B, 1, 3
caBench-to-Bedside. *See* caB2B, *See* caB2B
CDC, 54
Concept code, 16
Creating Experiments, 46
Custom Data Category
  creating, 54
  definition, 54
Data
  filtering, 46, 50, 51
  visualizing, 46, 50, 56
Data List, 13, 42, 46
  Saving, 44
Define Search Results View, 13, 24
Editing filters, 52
Executing saved query, 32
Experiment Details page, 49
  Operations panel, 50
  Visualization panel, 50
Experiments
  creating, 46
  definition, 46
  opening, 46, 49
  viewing, 48
Filters
  editing, 52
  enumerated value, 69
  pattern, 69
  range, 68
  removing, 53
  saving, 46, 50, 53
  types, 68
Limits
  connecting, 20
  DAG, 18
  defining, 13, 16, 19, 25
  definition, 12
  deleting, 20
  editing, 19
  information panel, 23
Login, 3, 4
My Data List Summary, 28, 30, 31, 32, 37
Opening Experiments, 46, 49
Path, 20
Plotting data series, 56
Projects, 46
Query
  definition, 12
  executing, 32
  saving, 13, 29
Query Wizard, 13, 46
Related data, 36
Removing filters, 53
Save query
  advantages, 29
  executing, 32
  non-parameterized, 30
Search categories, 60
Spreadsheet view, 49
View Search Results, 34
Viewing Experiments, 48
Vizualise data, 46, 50, 56
Vizualize data
  bar chart, 56, 57
  line chart, 56, 58
  scatter plot, 56, 58